



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 111165

TO: Michael Borin
Location: cm1/12a01/12d01
Art Unit: 1631
Monday, January 05, 2004

Case Serial Number: 09/930169

From: Toby Port
Location: Biotech-Chem Library
CM1-6A04
Phone: 308-3534

toby.port@uspto.gov

Search Notes

Dear Examiner Borin,

Here are the results of your search.
Please feel free to contact me if you have any questions.

Toby Port

THIS PAGE BLANK (USPTO)

O'Bry n, Barbara

11/65

Fr m: Borin, Michael
Sent: Tuesday, December 30, 2003 3:14 PM
T : STIC-Biotech/ChemLib
Subject: Search request: 09/930169

Examiner: M.Borin
CM1 12A01
AU: 1631; Mailbox 12D01

Tel.: 305-4506

RE: 09/930169; p43 fragment

Please conduct search of polypeptide SEQ ID 1,2 against the commercial and interference protein databases.

Thank you

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:17:12 ; Search time 11.5294 Seconds
(without alignments)
599.590 Million cell updates/sec

Title: US-09-930-169-1
Perfect score: 709
Sequence: 1 MANNDAVLKRLKQKGAADQ.....EKGKKEKQKQSIAGSADS 147

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_41.*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	709	100.0	312	1	MCAL_HUMAN
2	548.5	77.4	559	1	MCAL_CRICK
3	509	71.8	310	1	MCAL_MOUSE
4	125.5	17.7	516	1	P54_ENTPC
5	117	16.5	662	1	TLPB_BACSU
6	116.5	16.4	1079	1	IT2P_SCHU
7	108.5	15.3	662	1	MCPE_BACSU
8	106.5	15.0	1093	1	TWFI_HUMAN
9	106	15.0	461	1	US45_LACLC
10	104	14.7	473	1	YVCE_BACSU
11	104	14.7	570	1	STIM_DROME
12	104	14.7	961	1	VDP_BOVIN
13	103.5	14.6	561	1	EZEA_BACGD
14	103	14.5	1130	1	Y117_CABEL
15	103	14.5	1169	1	SMC_METUA
16	103	14.5	2230	1	GOG4_HUMAN
17	102	14.4	886	1	RA50_SITAC
18	102	14.4	1220	1	IT2P_HUMAN
19	101.5	14.3	539	1	MY33_HYDAT
20	101.5	14.3	1068	1	H1EP_MOUSE
21	101	14.2	956	1	KF5C_MOUSE
22	101	14.2	957	1	KF5C_HUMAN
23	101	14.2	1938	1	MYSS_CHICK
24	100.5	14.2	1290	1	C190_DROME
25	100	14.1	1243	1	SMC4_MICRA
26	99.5	14.0	886	1	RA50_AICRFU
27	99.5	14.0	2116	1	MY32_DICDI
28	99	14.0	1940	1	MY33_CHICK
29	98.5	13.9	407	1	Y173_AQUAE
30	98.5	13.9	879	1	RA50_SITAC
31	98	13.8	919	1	INCE_HUMAN
32	98	13.8	1197	1	C1NG_HUMAN
33	97.5	13.8	553	1	Y0IK_ECOLI

34	97.5	13.8	727	1	MFPI_ARATH	G91w65 arabidopsis
35	97	13.7	1605	1	RRBI_MOUSE	G99p15 mus musculus
36	97	13.7	2434	1	YCF1_OENHO	G9mth5 oenothera h
37	96.5	13.6	843	1	YMS1_YENST	G05050 saccharomyc
38	96	13.5	856	1	CLPB_HELPY	G71404 heliobacte
39	96	13.5	1526	1	MY32_SCHU	G9u516 schizosacch
40	96	13.5	1534	1	RRBI_CANFA	G28238 canis famli
41	95.5	13.5	793	1	CALD_HUMAN	G00562 homo sapien
42	95.5	13.5	1084	1	MYSS_RABIT	P02562 oryctolagus
43	95.5	13.5	1447	1	SGS1_YENST	P35187 saccharomyc
44	95.5	13.5	3214	1	BPAL_HUMAN	G03001 homo sapien
45	95	13.4	579	1	G160_HUMAN	G08378 homo sapien

ALIGNMENTS

RESULT 1
MCAL_HUMAN STANDARD; PRT; 312 AA.
ID MCAL_HUMAN
AC Q12904; Q96C09; PRT; 312 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Multienzyme complex auxiliary component p43 [Contains:
DE Endothelial-monocyte activating polypeptide II (EMAP-II) (small
DE inducible cytokine subfamily B member 1)].
GN SCF1 OR EMAP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95014290; PubMed=7929199;
RA Kao J., Houck K., Fan Y., Haehnel I., Libutti S.K., Kayton M.L.,
RA Grikscheit T., Chabot J., Nowygrod R., Greenberg S., Kuang W.J.,
RA Leung D.W., Hayward J.R., Kiesel W., Heath M., Brett J., Stern D.M.;
RT "Characterization of a novel tumor-derived cytokine. Endothelial-
RT monocyte activating polypeptide II,"
RL J. Biol. Chem. 269:25106-25119(1994)
RP SEQUENCE FROM N.A.
RC TTSUR-Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marisina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stetson M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulvaney S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buttefield Y.S.N., Krzyzanski M.I., Skalek U., Smalins D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Maria W.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences,"
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SIMILARITY: Contains 1 tRNA-binding domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

THIS PAGE BLANK (USPTO)

or send an email to license@ib-sib.ch).

CC or send an email to license@ib-sib.ch).

CC EMBL; U10117; AAA62202.1; -

DR EMBL; BC014051; AAA14051.1; -

DR PDB; 1E7Z; 06-FEB-01.

DR PDB; 1EUJ; 06-SEP-00.

DR PDB; 1FL0; 07-FEB-01.

DR Genew; HGNC:10648; SCYB1.

DR MIM; 603605; -

DR GO; GO:0005615; C:extracellular space; TAS.

DR GO; GO:0005125; F:cytokine activity; TAS.

DR GO; GO:0000049; F:cytokine activity; TAS.

DR GO; GO:0006418; P:amino acid activation; TAS.

DR GO; GO:0006935; P:chemotaxis; TAS.

DR GO; GO:0006954; P:inflammatory response; TAS.

DR InterPro; IPR002547; tRNA bind.

DR Pfam; PF01588; tRNA bind.

DR PROSITE; PIRSF005381; EMAP11; 1.

DR PROSITE; PSS0886; TRBD; 1.

KM Protein biosynthesis; RNA-binding; tRNA-binding; Cytokine;

KM 3D-structure.

FT PROPEP 1 146

FT CHAIN 147 312

FT DOMAIN 151 252

FT CONFLICT 79 79

FT COMPLET A -> P (IN REF. 2).

SO SEQUENCE 312 AA; 34326 MW; 946310A0216F7587 CRC64;

Query Match Best Local Similarity 100.0%; Score 709; DB 1; Length 312; Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANDAVLRLKLEOKGAEADQIIIEYLRQVSLKEKALIQATLRREKKLRVENAKLKEIE 60

DB 1 MANDAVLRLKLEOKGAEADQIIIEYLRQVSLKEKALIQATLRREKKLRVENAKLKEIE 60

QY 61 ELKQELIOAEIQNGVQKQIAPFSGTGPLHANSVSNVIOSTAVTTVSSGTKEQIKGTGDE 120

DB 61 ELKQELIOAEIQNGVQKQIAPFSGTGPLHANSVSNVIOSTAVTTVSSGTKEQIKGTGDE 120

QY 121 KKAKEKIEKGEKKEKKQKQSIAGSADS 147

DB 121 KKAKEKIEKGEKKEKKQKQSIAGSADS 147

RESULT 2

MCAL_CRIGR STANDARD; PRT; 359 AA.

AC 054873; 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Multisynthetase complex auxiliary component p43 [contains: Endothelial-monocyte activating polypeptide II (EMAP-II) (Small inducible cytokine subfamily B member 1)].

GN SCYB1.

OS Cricetulus griseus (Chinese hamster).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.

OC NCBI_TaxID=10029;

CC NCBI_TaxID=10029;

CC SEQUENCE FROM N.A.

CC MEDLINE=98070438; PubMed=9405472;

RA Quevillon S., Agou F., Robinson J.-C., Mirande M.;

RT "The p43 component of the mammalian multi-synthetase complex is likely to be the precursor of the endothelial monocyte-activating polypeptide II cytokine."

RT J. Biol. Chem. 272:32573-32579(1997).

CC -1- SUBUNIT: COMPONENT OF THE MULTISYNTHETASE COMPLEX WHICH IS COMPRISED OF A BIFUNCTIONAL GLUTAMYL-PROLYL-TRNA SYNTHETASE, THE MONOSPECIFIC ISOLEUCYL, LEUCYL, GLUTAMINYL, METHIONYL, LYSYL, ARGINYL, AND ASPARYL-TRNA SYNTHETASES AS WELL AS THREE AUXILIARY

PROTEINS, P18, P48 AND P43.

-1- SIMILARITY: Contains 1 tRNA-binding domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@ib-sib.ch).

CC EMBL; AF021800; AA95207.1; -

DR InterPro; IPR002547; tRNA bind.

DR Pfam; PF01588; tRNA bind.

DR PIRSF; PIRSF005381; EMAP11; 1.

DR PROSITE; PSS0886; TRBD; 1.

KM Protein biosynthesis; RNA-binding; tRNA-binding; Cytokine.

FT DOMAIN 198 299

SO SEQUENCE 359 AA; 39601 MW; 4D868D1B65D72C23 CRC64;

Query Match Best Local Similarity 77.4%; Score 548.5; DB 1; Length 359; Matches 119; Conservative 10; Mismatches 17; Indels 3; Gaps 2;

QY 1 MANDAVLRLKLEOKGAEADQIIIEYLRQVSLKEKALIQATLRREKKLRVENAKLKEIE 60

DB 47 MANDAVLRLKLEOKGAEADQIIIEYLRQVSLKEKALIQATLRREKKLRVENAKLKEIE 106

QY 61 ELKQELIOAEIQNGVQKQIAPFSGTGPLHANSVSNVIOSTAVTTVSSGTKEQIKGTG 118

DB 107 ELKQELIOAEIQNGVQKQIAPFSGTGPLHANSVSNVIOSTAVTTVSSGTKEQIKGTG 165

QY 119 DEKKAKEKIEKGEKKEKKQKQSIAGSADS 147

DB 166 DEKKAKEKIEKGEKKEKKQKQSIAGSADS 194

RESULT 3

MCAL_MOUSE STANDARD; PRT; 310 AA.

AC P31230; Q06659; 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Multisynthetase complex auxiliary component p43 [contains: Endothelial-monocyte activating polypeptide II (EMAP-II) (Small inducible cytokine subfamily B member 1)].

GN SCYB1 OR EMAP2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI_TaxID=10090;

CC NCBI_TaxID=10090;

CC SEQUENCE FROM N.A.

CC MEDLINE=95014290; PubMed=7929199;

RA Kao J., Houck K., Fan Y., Haehnle I., Libutti S.K., Kayton M.L., Grikscheit T., Chabot J., Nowygrad R., Greenberg S., Xiang W.J., Leung D.W., Hayward J.R., Kistel W., Heath N., Brett J., Stern D.M.;

RT "Characterization of a novel tumor-derived cytokine, Endothelial-monocyte activating polypeptide II."

RT J. Biol. Chem. 269:25106-25119(1994).

CC NCBI_TaxID=10090;

CC SEQUENCE FROM N.A.

CC MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Parner A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavani T.L., Scheetz T.E., Brownstein W.J., Udell T.B., Toshiyuki S., Carninci P., Prange C., Rana S.S., Loggellano N.A., Peters G.J., Abramson R.D., Miliady S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - protein search, using SW model

Run on: January 2, 2004, 18:16:37 ; Search time 43.8118 Seconds
(without alignments)
532.569 Million cell updates/sec

Title: US-09-930-169-1

Perfect score: 709
Sequence: 1 MANDAVAKRLKQKGAEDQ.....EKKGKKEKKQKQSIAGSADS 147

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.GeneSeq_19Jun03:*

- 1: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:*
- 23: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:*
- 24: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	709	100.0	312	23	ABB07442
2	709	100.0	328	22	AAB63244
3	509	71.8	310	16	AAR72577
4	119.5	16.9	294	22	ABB59210
5	111.5	15.7	1047	22	AAW76507
6	111.5	15.7	1047	22	AAW79491
7	109	15.4	461	12	AAK14150
8	108	15.2	456	23	ABB55584
9	108	15.2	461	12	AAK14530

10	106.5	15.0	1093	14	AAK42818
11	105.5	14.9	1122	24	ABU70418
12	105.5	14.9	1193	22	AAW25602
13	105.5	14.9	1193	23	ABG61824
14	105.5	14.9	1194	22	AAU32407
15	105.5	14.9	1833	22	ABW71141
16	105.5	14.9	1907	23	ABB82127
17	105.5	14.9	5533	22	ABB65772
18	105.5	14.9	5560	22	ABW71160
19	105	14.8	1975	22	ABB62094
20	104	14.7	516	21	AAE52007
21	104	14.7	516	21	AAE51636
22	104	14.7	1066	22	ABB64251
23	103	14.5	286	19	AAW98376
24	103	14.5	941	22	ABU3072
25	103	14.5	951	22	ABU3070
26	103	14.5	953	22	ABU3069
27	103	14.5	953	22	ABU3071
28	103	14.5	961	22	ABU53077
29	103	14.5	2096	21	AAW41592
30	103	14.5	2230	24	ABU7445
31	102	14.4	290	20	AAW97293
32	102	14.4	431	20	AAV9137
33	101.5	14.3	271	19	AAW42078
34	100.5	14.2	419	20	AAV49149
35	100.5	14.2	433	20	AAV49245
36	100.5	14.2	433	20	AAV32184
37	100.5	14.2	672	21	ABB21233
38	100.5	14.2	1075	23	ABB77436
39	100.5	14.2	1690	22	ABB61144
40	100.5	14.2	1690	22	ABB61173
41	100.5	14.2	1978	20	AAV07032
42	100	14.1	388	21	AAW21232
43	100	14.1	589	20	AAV43392
44	100	14.1	715	23	ABU05544
45	100	14.1	929	18	AAW14593

ALIGNMENTS

RESULT 1	ABB07442	ABW7442
ID	ABB07442	standard; Protein; 312 AA.
AC	ABB07442;	
XX		
DT	23-APR-2002	(first entry)
XX		
DE	Human p43 polypeptide.	
XX		
KW	Human; p43; cytosolic; anti-tumour; anti-angiogenic; EMAP II;	
KW	cytokine; chemotaxis.	
OS	Homo sapiens.	
XX		
PH	Key	Location/Qualifiers
FT	Domain	1..146
FT	Domain	/note= "N-terminal domain"
FT	Domain	147..312
FT	Domain	/note= "C-terminal domain"
FT	Region	10..21
FT	Region	/note= "beta1 strand"
FT	Region	28..34
FT	Region	/note= "beta2 strand"
FT	Region	40..46
FT	Region	/note= "beta3 strand"
FT	Region	53..56
FT	Region	/note= "alpha 1 helix"
FT	Region	59..66
FT	Region	/note= "beta4 strand"
FT	Region	70..72
FT	Region	/note= "beta5 strand"

TMF. Homo sapiens
Human adipocyte Se
Human protein sequ
Prostate cancer-as
Novel human secret
Drosophila melanog
Human TANGO 130 po
Drosophila melanog
Drosophila melanog
M. jamaeichi M08
M. jamaeichi M08
Drosophila melanog
H. pylori GHPO 131
Intracellular traf
Intracellular traf
Intracellular traf
Intracellular traf
Human ORF1356
Protein different
An annexin binding
Amino acid sequenc
The amino acid seq
Amino acid sequenc
N-terminal region
N-terminal choline
Corn MFPI. Zea ma
Human tumour marke
Drosophila melanog
Drosophila melanog
Breast cancer asso
Soybean MFPI. Gly
BpC alpha-helix c
Breast cancer-asso
Streptococcus pneu

FT	Region	75..77
FT	/note= "beta6i strand"	
FT	Region	79..85
FT	/note= "beta7 strand"	
FT	Region	90..92
FT	/note= "beta8 strand"	
FT	Region	103..106
FT	/note= "beta9 strand"	
FT	Region	119..123
FT	/note= "alpha 2 helix"	
FT	Region	124..130
FT	/note= "alpha 3 helix"	
FT	Region	132..134
FT	/note= "beta10 strand"	
FT	Region	140..142
FT	/note= "beta1i strand"	
XX		
PN	MOZ00195927-AI.	
PD	20-DEC-2001.	
PF	14-JUN-2000; 2000WO-KR00630.	
PR	14-JUN-2000; 2000WO-KR00630.	
XX		
PA	(IMAG-) IMAGEGENE CO LTD.	
PI	Kim S, Ko Y, Kim YS, Jo YJ;	
XX		
DR	WPI; 2002-098017/13.	
DR	N-PsDB; ABA94640.	
XX		
PT	Novel anti-tumour and anti-angiogenic agent of p43 comprises N-terminal domain and C-terminal domain containing eleven beta-strands forming a structural core and three flanking alpha-helices -	
PS	Claim 1; Fig 1; 35pp; English.	
XX		
CC	The invention provides an anti-tumour and anti-angiogenic agent of p43 consisting of two domains, the N-terminal domain (146 amino acids) and C-terminal domain (166 amino acids) containing 11 beta-strands forming a structural core and 3 flanking alpha-helices. p43 is useful as an anti-tumour and anti-angiogenic agent. p43 and its C-terminal cytokine domain (EMAP II) induce regression of fibrosarcoma in immunocompromised mouse while its N-terminal domain does not. p43 is a potent cytokine as determined by the induction of tumour necrosis factor-alpha (TNF-alpha), interleukin-6 (IL-6), IL-8 and matrix metalloproteinase-9 or by its activity of chemotaxis. The present sequence represents the human p43 polypeptide.	
CC		
XX		
SO	Sequence 312 AA;	
Query Match	100.0%; Score 709; DB 23; Length 312;	
Best Local Similarity	100.0%; Pred. No. 4.3e-57;	
Matches 147; Conservative	0; Mismatches 0; Indels 0; Gaps 0	
Dd	1 MANNDVAVLKRLLEQKGAADQIIIEYKKOVSILKEKAILIQATLRBEKKLREVENAKLKEIE 60	
Qy	MANNDVAVLKRLLEQKGAADQIIIEYKKOVSILKEKAILIQATLRBEKKLREVENAKLKEIE 60	
Dd	1 MANDVAIVKLRLQKGAADQIIIEYKKOVSILKEKAILIQATLRBEKKLREVENAKLKEIE 60	
Qy	ELKQELIOAEITONGVKQIAFPSGTPULHANSWSENVIOSTAVTVSSGTKEQIKGGTGE 120	
Dd	61 ELKQELIOAEITONGVKQIAFPSGTPULHANSWSENVIOSTAVTVSSGTKEQIKGGTGE 120	
Qy	121 KKAKEIKIEKKGEKKKKKOOSTIGSDS 147	
Dd	121 KKAKEIKIEKKGEKKKKKOOSTIGSDS 147	
RESULT 2		
AAB63244		
ID	AAB63244 standard; Protein; 328 AA.	
XX		

AC	AAB63244;
XX	
DT	26-MAR-2001 (first entry)
DE	
XX	Human breast cancer associated antigen protein sequence SEQ ID NO:606.
KW	Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KX	cancer associated antigen; cytostatic; cancer vaccine.
OS	Homo sapiens.
PV	WO20073801-A2.
PD	07-DEC-2000.
PF	26-MAY-2000; 2000MO-USI4749.
PR	28-MAY-1999; 99US-0136526.
PT	10-SEP-1999; 99US-0153454.
PA	(LUDM-) LUDMWG INST CANCER RES.
PI	Obata Y;
WP	WIPO; 2001-025274/03.
Nucleic acid encoding breast, gastric and prostate cancer associated antigen precursors, useful for diagnosing and treating a condition characterized by expression of an abnormal amount of a protein, e.g. cancer - Example 1; Page 478; 79pp; English.	
AAE22422 to AAF22626, AAF22627 to AAF22773 and AAE22774 to AAF23014 represent nucleotide sequences encoding human breast, gastric and prostate cancer associated antigen precursors (CAAP) respectively. AAB63332 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970 represent human breast, gastric and prostate CAAP protein sequence respectively. CAAPs have cytosolic activity and can be used in the production of cancer vaccines. The human CAAP proteins, peptides, nucleic acids or anti-CAAP antibodies are useful for diagnosing and treating a condition characterised by expression of an abnormal amount of a protein, e.g. cancer.	
CX CC CC CX	Sequence 328 AA; SQ
Query Match	Score 709; DB 22; Length 328; Best Local Similarity 100.0%; Prd. No. 4,6e-57; Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DQ	1 MANNDIVLKRLKLGKAADQLIEYLRKKOVSLLKEKAILQATLRREKLIVENAKLKEIE 60 DB 17 MANNDAVLKRLEGGAGEAHQITETLVKQQVSLRKESKALIQATVRREKCLRVENAKLKEIE 76
OY	61 ELKOELLQAELIONGVKQIAFPSSGTPLHANSWSENVIOSTAVTVTSSGTKQIKGTIDE 120 DB 77 ELKOELLQAELIONGVKQIAFPSSGTPLHANSWSENVIOSTAVTVTSSGTKQIKGTIDE 136
OY	121 KKAKEKTEKKGEKKKKQKSASGS 147 DB 137 KKAKEKTEKKGEKKKKQKSASGS 163
RESULT 3	
ID	AAF2577 standard; Protein: 310 AA.
XX	AAF2577;
AC	
XX	25-MAR-2003 (updated)
DT	29-SEP-1995 (first entry)
XX	Mouse EMAPHI

```

XX EMAP11; endothelial monocyte activating polypeptide II; chemotaxis;
KW inflammation; tissue factor; tumor; cancer; therapy; metha;
KM sarcoma.
XX
XX Mus musculus.
OS
XX MO9509180-A1.
PN
XX
XX 06-APR-1995.
PD
XX
XX 29-SEP-1994; 94MO-US11085.
PF
XX 29-SEP-1993; 93US-0129456.
PR
XX (UYCO ) UNIV COLUMBIA NEW YORK.
PA
XX
XX Claus M, Kao J, Kayton M, Libutti SK, Stern DM;
PI
XX WPI; 1995-147389/19.
DR
XX N-PSDB; AAQ86718.
DR
XX
XX New endothelial monocyte activating polypeptide II - induces
PT chemotaxis, inflammation and tissue factor, useful for treating
PT tumours, also related antibodies, DNA and active fragments
XX
XX PS
XX Disclosure; Fig.4; 180pp; English.
XX
XX A mouse metha sarcoma CDNA library was screened with a probe
CC based on the N-terminal sequence of mouse EMAP11. Overlapping
CC clones were combined to obtain a contiguous full-length sequence
CC (given in AAQ86718) encoding a 33 kDa protein (AAR72577).
CC Recombinant EMAP11 was expressed in E. coli.
CC (Updated on 25-MAR-2003 to correct PN field.)
CC
XX
XX Sequence 310 AA;
SQ
Query Match 71.8%; Score 509; DB 16; Length 310;
Best Local Similarity 76.4%; Pred. No. 1.1e-38;
Matches 113; Conservative 8; Mismatches 23; Indels 4; Gaps 3;
QY 1 MANDAVLKRLKQKGAENDQIIIEYKQVSLKKEKALIQATLREKKLRVENAKLKKEIE 60
DB 1 MATNDVAVLKRLKQKGAENDQIIIEYKQVSLKKEKALIQATLREKKLRVENAKLKKEIE 60
QY 61 ELKQELIOAEIONGVKQIAFPSPGPIHANSWSENVIOSTAV-TTVSSGTKEQIKGCTG 119
DB 61 ELKQELILAEITHNGVEQVRRLSTPLQNTCTASBSVQSPVATYASPATKEQIK--AGE 118
QY 120 EKKAKERIEKKKKEKKKQOSIAGSADS 147
DB 119 EKKYKEKEKEKKEKKE--KQSAASSTDS 145
RESULT 4
ABBS9210
ID ABBS9210 standard; Protein; 294 AA.
XX
XX
XX ABB59210;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 4422.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX

```

```

PF 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PB CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
DR
XX N-PSDB; ABL03313.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX PS
XX Disclosure; SEQ ID NO 4422; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
XX
XX Sequence 294 AA;
SQ
Query Match 16.9%; Score 119.5; DB 22; Length 294;
Best Local Similarity 24.8%; Pred. No. 0.007;
Matches 37; Conservative 33; Mismatches 52; Indels 27; Gaps 4;
QY 6 AVLRKLEKQKGAENDQIIIEYKQVSLKKEKALIQATLREKKLRVENAKLKKEIEELKOE 65
DB 2 ADLQGISNNRERAKALINSIEAEISGIGQ----QLVERQKQELKENNALAKEVYALAQ 57
QY 66 LIQAEIONGVKQIAFP-----SGTPIHANSWSENVIOSTAVTTVSSGTKEQIKGCTG 118
DB 58 LVQELNLNGKQKQIPVPGARGFCTSAAPV---VMPAEGPATAPAPA----- 102
QY 119 DEKAKERIEKKKKEKKKQOSIAGSADS 147
DB 103 -PKPAKEPEKKEKKEKPAEKPAPAEPA 130
RESULT 5
AAM78507
ID AAM78507 standard; Protein; 1047 AA.
XX
XX
XX AAM78507;
AC
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human protein SEQ ID NO 1169.
DE
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukemia;
KW nervous system disorder; arthritis; inflammation.
XX
XX OS
XX Homo sapiens.
XX
XX WO200157190-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 05-FEB-2001; 2001WO-US04098.
PF
XX
XX 03-FEB-2000; 2000US-0496914.
PR
XX

```

PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
DR N-PSDB; AAK51640.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 3407-3409; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoietic regulatory
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX Sequence 1047 AA;
SQ
Query Match 15.7%; Score 111.5; DB 22; Length 1047;
Best Local Similarity 23.0%; Pred. No. 0.2;
Matches 38; Conservative 43; Mismatches 49; Indels 35; Gaps 5;
QY 8 IKRLEQKGAADQIIEYLKQOVSILKEKAILQATLREKKLRVNAKLEIELEKQELI 67
Db 736 LEKLDVYRQQAQAIIFLEKQISLAERKML-----DYERLQRAEAGQGEVSLREKLL 769
QY 68 QAEIQ-NGVKQIAFPSTPLHANSWSENVIOS-----TAVTVSSG 108
Db 790 VAENRLQAVFALCSSQHTHIESNDISEETIRKETEVEGLQDKLNKRDKEVTALTSQTEM 849
QY 109 TKEQI-----KGGTGDEK-----KAKEKIEKKKKEKKQOOSIAG 143
Db 850 LRAQVSALSKCKSGEKKVDALLKEKRRLAELETVSRKTHDASG 894
RESULT 6
AAM79491
ID AAM79491 standard; Protein; 1047 AA.
XX
XX AAM79491;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human protein SEQ ID NO 3137.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KM vaccine; peptide therapy; stem cell growth factor; haematopoietic;
KM tissue growth factor; immunomodulatory; cancer; leukaemia;
KM nervous system disorder; arthritis; inflammation.
XX
XX Homo sapiens.
XX

PN W0200157190-A2.
XX
XX 09-AUG-2001.
PD
XX
XX 05-FEB-2001; 2001MO-US04098.
PF
XX
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
DR N-PSDB; AAK52624.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 262; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoietic regulatory
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX Sequence 1047 AA;
SQ
Query Match 15.7%; Score 111.5; DB 22; Length 1047;
Best Local Similarity 23.0%; Pred. No. 0.2;
Matches 38; Conservative 43; Mismatches 49; Indels 35; Gaps 5;
QY 8 IKRLEQKGAADQIIEYLKQOVSILKEKAILQATLREKKLRVNAKLEIELEKQELI 67
Db 736 LEKLDVYRQQAQAIIFLEKQISLAERKML-----DYERLQRAEAGQGEVSLREKLL 769
QY 68 QAEIQ-NGVKQIAFPSTPLHANSWSENVIOS-----TAVTVSSG 108
Db 790 VAENRLQAVFALCSSQHTHIESNDISEETIRKETEVEGLQDKLNKRDKEVTALTSQTEM 849
QY 109 TKEQI-----KGGTGDEK-----KAKEKIEKKKKEKKQOOSIAG 143
Db 850 LRAQVSALSKCKSGEKKVDALLKEKRRLAELETVSRKTHDASG 894
RESULT 7
AAR14150
ID AAR14150 standard; Protein; 461 AA.
XX
XX AAR14150;
XX
XX 25-MAR-2003 (updated)
DT 09-DEC-1991 (first entry)
XX
XX MSP encoded by pUCRS (DSM 5803).
DE

[illegible]

OS	Lactococcus lactis IL1403.
XX	
PN	FR2807446-A1.
XX	
PD	12-OCT-2001.
XX	
PF	11-APR-2000; 2000FR-0004630.
XX	
PR	11-APR-2000; 2000FR-0004630.
XX	
PA	(INRG) INRA INST NAT RECH AGRONOMIQUE.
XX	
PI	Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX	
DR	WPI; 2002-043418/06.
XX	
PT	New nucleotide sequence useful in the identification or Lactococcus
XX	
PT	lactis and related species -
XX	
PS	Claim 6; SEQ ID NO 2286; 2504bp; French.
XX	
CC	The present invention is related to a Lactococcus lactis nucleotide
XX	
CC	sequence (AB990521) and related proteins (AB95300-AB955621). The
XX	
CC	nucleic acid sequence is useful in the detection and/or amplification of
XX	
CC	nucleic acid sequence, particularly to identify Lactococcus lactis or
XX	
CC	related species. The proteins of the invention are useful for the
XX	
CC	biosynthesis or biodegradation of a composition of interest. The
XX	
CC	invention helps research in lactic bacteria, particularly useful in the
XX	
CC	production of yogurt and cheese.
XX	
CC	Note: The sequence data for this patent is based on equivalent patent
XX	
CC	WO200177334 (published 18-Oct-2001) which is available in electronic
XX	
CC	format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 456 AA;
Query Match	15.2%; Score 108; DB 23; Length 456;
Best Local Similarity	26.6%; Pred. No. 0.14;
Matches 42; Conservative 33; Mismatches 65; Indels 18; Gaps 6;	
OY	1 MANDDAVKKRLKRGAGADQIIEYLKQOVSLKEKALIQATLR-----EKKLRVEN 52
DB	33 IAKDPAITSSAQSAKAQAQAVDSLQSVDSLQOK--QASTKQIAKISEBAKLNAOI 89
OY	53 AKLKKEIELKOEHLQAEIOAIBIONGVQAIAPPSTPLHANSMVSENYIOS-TAVTVSSGTKE 111
DB	90 ATLNESIALF-RTKTLEAQRASQAQVNSSTNTNMDAVNMSKSILTDVIQKVTAIATVSSANKO 148
OY	112 QIKGDTGDDEKKAKERIEKGKEREKOQ--SIAGSADS 147
DB	149 MLE---QQEKREKELSGKSETVKKNVQNFVLSGLSDLS 183
RESULT 9	
ID	AARI4530 standard; Protein; 461 AA.
XX	
AC	AARI4530;
XX	
DT	25-MAR-2003 (updated)
XX	
DT	28-JAN-1992 (first entry)
XX	
DE	Usp45 protein.
XX	
KM	Chymosin; alpha-amylase; PNZ1011; enzyme.
XX	
OS	Lactococcus lactis spp. lactis MG1363.
XX	
FH	Key Location/Qualifiers
FT	Peptide 1..27
FT	/label= sig_peptide
FT	Protein 28..461
FT	/label= mature_protein
XX	

PN NL9000753-A.
 XX 16-OCT-1991.
 PD 30-MAR-1990; 90NL-0000753.
 XX 30-MAR-1990; 90NL-0000753.
 PR 30-MAR-1990; 90NL-0000753.
 XX (NEZU-) NEDERLANDS INST ZUI.
 PA WPI; 1991-330695/45.
 XX N-PSDB; AAQ14501.
 DR Lactococcus lactis DNA fragments - contg. extracellular protein
 XX signal peptide sequence
 PT Lactococcus lactis DNA fragments - contg. extracellular protein
 XX signal peptide sequence
 PS Disclosure; Fig 4; 20pp; Dutch.
 XX The protein has a mol.wt. of 60 kD and is encoded by the KpnI/ClaI
 CC fragment from pNZ1011. Lactococci transformed with the usp45 gene may
 CC be used to produce extracellular heterologous enzymes of use in the food
 CC industry, e.g. chymosin or alpha-amylase.
 CC See also AAQ14502.
 CC (Updated on 25-MAR-2003 to correct PF field.)
 CC
 SQ Sequence 461 AA;
 Query Match 15.2%; Score 108; DB 12; Length 461;
 Best Local Similarity 26.5%; Pred. No. 0.15; Mismatches 69; Indels 12; Gaps 5;
 Matches 41; Conservative 33;
 QY 1 MANNDVAVLKRLKLEQGAADQIIEYLKQVSLKE-----KAIQATLREKKLRVENAKL 55
 DB 33 IAKQDARISAGAKAQAGQVDSLOSKVSLQKQSTYAKIASEKAKALNAQIATL 92
 QY 56 KKEIEIKQELIQAEIQNGVKQIAFPSTPLHANSVSENVIGS-TAVTVSSGTRKQIK 114
 DB 93 NESIKK-RTKTLERQARSAGVNSATNYMDAVVNSKSLTDVIGVTAIVSSANKQMLE 151
 QY 115 GGTGDEKAKKEIEKKEKKEKQKQ--SINGSADS 147
 DB 152 ---QOEKQKELSQKSETVKKNYNQFVLSQSDS 183
 RESULT 10
 AAR42818
 ID AAR42818 standard; Protein; 1093 AA.
 XX
 XX AAR42818;
 AC
 XX 25-MAR-2003 (updated)
 DT 27-APR-1994 (first entry)
 XX
 XX TMF.
 DE
 XX TATA modulating factor; TMF; transcription; TATA box; promoter; HIV-1;
 KW human immunodeficiency virus-1; short arm; human chromosome 3; p12-p21;
 KW translocation; cancer.
 XX
 XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT Region 437..850
 FT /label= TATA binding region
 FT Region 769..777
 FT /note= "Ubiquitin-mediated protein degradation
 FT consensus sequence homology region"
 FT 454..614
 FT /note= "Region with Leucine zipper secondary
 FT structure"
 FT Region 986..1069
 FT /note= "Region with Leucine zipper secondary
 FT structure"

FT Region 1070..1078
 FT /note= "Ubiquitin-mediated protein degradation
 FT consensus sequence homology region"
 XX
 XX MO9320106-A1.
 PN 14-OCT-1993.
 PD 31-MAR-1993; 93WO-US03077.
 XX 02-APR-1992; 92US-0862025.
 PR (TEXA) UNIV TEXAS SYSTEM.
 XX Gaynor RB, Wu F;
 XX WPI; 1993-336836/42.
 DR N-PSDB; AAQ49397.
 XX New protein cellular factor - capable of binding double stranded
 PT HIV-1 tata region and activating gene expression of HIV-LTR
 XX Claim 2; Fig 1; 75pp; English.
 PS This sequence represents TATA modulating factor (TMF). TMF is a
 CC protein of mol. wt. 123-130 kD which activates transcription in most
 CC genes, esp. in human immunodeficiency virus-1 (HIV-1) by binding to
 CC the TATA box region of the promoter. TMF is encoded by the short
 CC arm of human chromosome 3 in the region p12-p21 which is often
 CC involved in translocations in patients having lung and other types
 CC of cancer.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC
 SQ Sequence 1093 AA;
 Query Match 15.0%; Score 106.5; DB 14; Length 1093;
 Best Local Similarity 22.4%; Pred. No. 0.62;
 Matches 37; Conservative 33; Mismatches 50; Indels 45; Gaps 5;
 QY 3 NNDVAVLKRLKLEQGAADQIIEYLKQVSLKE-----KAIQATLREKKLRVENAKL 57
 DB 571 HNSNIITKRLRAVDKENENNAKLNKVKLEELQHLKQVDSLOKKEVEKQRENIKKLNS 630
 QY 58 EIEELKQEL-----IQAEIQNGVKQIAFPSTPLHANSVSENVIGSTA 101
 DB 631 MYRQEKDGLRQVMDLEKKRSIQALDSAYKEL-----TDLIKANAADSEAGQEA 685
 QY 102 VTVSSGTRKQIKGTGDEKAKK-----IEKKKKEKKQOSIA 142
 DB 686 LSR-----EMKAKKELSALEKQAEARQOQETLA 715
 RESULT 11
 ABU70418
 ID ABU70418 standard; Protein; 1122 AA.
 XX
 XX ABU70418;
 AC
 XX 10-JUN-2003 (first entry)
 DT
 XX Human adipocyte Selected Interacting domain; SID, #49.
 DE
 XX Human; prey; adipocyte; SID; selected interacting domain;
 KW anorectic; antidiabetic; protein-protein interaction; diabetes;
 KW yeast 2-hybrid assay; metabolic disorder; obesity.
 XX
 XX Homo sapiens.
 OS
 XX
 PN WO200286122-A2.
 PD 31-OCT-2002.
 XX
 XX 14-MAR-2002; 2002WO-EP03768.

XX 14-MAR-2001; 2001US-275734P.
 PR (HYBR-) HYBRIGENICS.
 PA Legrain P, Daviet L;
 XX WPI; 2003-103412/09.
 DR N-PSDB; ACN56962.
 XX
 PT New complex between two interacting proteins in adipocyte cells, useful
 PT for identifying selected interacting domains that modulate protein
 PT interactions, or for preventing or treating metabolic disorders such as
 PT obesity or diabetes
 PS
 PS Claim 6; Page 127-128; 382pp; English.
 XX
 CC The invention relates to a complex between two interacting proteins in
 CC adipocyte cells, given in the specification. The proteins are identified
 CC by selecting a bait protein from a known adipocyte marker and then
 CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by
 CC members of an adipocyte cDNA library. The proteins are designated SID
 CC (RTM) (selected interacting domains) proteins. Also included are a
 CC polynucleotide encoding a polypeptide in the adipocyte cells, a
 CC recombinant host cell expressing at least one of the interacting
 CC polypeptides of the complex, selecting a modulating compound in adipocyte
 CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid
 CC sequences given in the specification (including its fragment or variant),
 CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
 CC given in the specification (including its fragment or variant), a vector
 CC comprising the SID (RTM) polynucleotide, a recombinant host cell
 CC comprising the vector, a protein chip comprising the polypeptides and
 CC a record comprising all or part of the data, listed in the specification.
 CC The complex, polypeptides, polynucleotides and compounds are
 CC useful for preventing or treating metabolic disorders such as obesity
 CC or diabetes. The polynucleotides are useful as probes or primers. The
 CC complex is particularly useful for identifying selected interacting
 CC domains (SID (RTM)) for screening drugs that modulate the protein
 CC interaction, thus exhibiting the therapeutic effect. The present
 CC sequence represents a SID (prey) protein of the invention.
 XX
 XX Sequence 1122 AA;
 SQ
 Query Match 14.9%; Score 105.5; DB 24; Length 1122;
 Best Local Similarity 26.1%; Pred. No. 0.79;
 Matches 41; Conservative 30; Mismatches 55; Indels 31; Gaps 6;
 QY 16 AADQIIYLYLKQVSLKKEK-----AIIQATLREK-----KLRYENAKIKKEIEBLK 63
 DB 520 SENKSIIEKLKDVISMNASSEFEVQIALNEAKLSEKVKSECHRVQENARLKKKKQLQ 579
 QY 64 QEL-----IQAEIONGVK-----QIAPSGTPIHANSVSNVIGSTAVTTVSSGT 109
 DB 580 QEIDWSKTLHAEISQIKSFPEKSKQDLVAL---THKDNINALTNCITQALNLECESSES 636
 QY 110 KEQIKGTGDEKKAKEIKKEKKEKKQOQSIAGSAD 146
 DB 637 EGQNKGGNDSPLANGEV--GGDRNKKKNQIKQMM 671
 RESULT 12
 AAM25602
 ID AAM25602 standard; Protein; 1193 AA.
 XX
 AC AAM25602;
 XX
 XX 16-OCT-2001 (first entry)
 DE Human protein sequence SEQ ID NO:1117.
 XX
 XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 KM antiinflammatory; antipneumatic; antidiabetic; immunosuppressive;
 KM antibacterial; endocrine; cardiant; central nervous system; virucide;

KM anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
 KM antiaggregant; haemostatic; vulnary; antilucer; osteopathic; eczema;
 KM dermatological; antiallergic; antiaesthetic; antidiabetic; cytostatic;
 KM neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KM immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KM antiaphyllactic; rheumatoid arthritis; septic shock; pancreatitis;
 KM cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KM genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KM thrombocytopenia; osteoporosis; severe combined immunodeficiency;
 KM allergic rhinitis; diabetes; multiple sclerosis; depression;
 KM Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KM neurological disorder.
 OS
 OS Homo sapiens.
 XX
 XX WO200153455-A2.
 XX
 XX 26-JUL-2001.
 XX
 XX 22-DEC-2000; 2000WO-US35017.
 XX
 XX 23-DEC-1999; 99US-0471275.
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 PI Tang YT, Liu C, Dymnac RT;
 DR WPI; 2001-457603/49.
 DR N-PSDB; AAM99543.
 XX
 PS
 PS Claim 20; Page 230; 1217pp; English.
 XX
 CC AAM99166 to AAM99904 encode the human proteins given in AAM25225 to
 CC AAM25963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnary;
 CC antifuric; osteopathic; dermatological; antiallergic; antiaesthetic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.
 XX
 XX Sequence 1193 AA;
 SQ
 Query Match 14.9%; Score 105.5; DB 22; Length 1193;
 Best Local Similarity 26.1%; Pred. No. 0.86;
 Matches 41; Conservative 30; Mismatches 55; Indels 31; Gaps 6;
 QY 16 AADQIIYLYLKQVSLKKEK-----AIIQATLREK-----KLRYENAKIKKEIEBLK 63
 DB 591 SENKSIIEKLKDVISMNASSEFEVQIALNEAKLSEKVKSECHRVQENARLKKKKQLQ 650
 QY 64 QEL-----IQAEIONGVK-----QIAPSGTPIHANSVSNVIGSTAVTTVSSGT 109
 DB 651 QEIDWSKTLHAEISQIKSFPEKSKQDLVAL---THKDNINALTNCITQALNLECESSES 707

QY 16 ABAADQIIEYLKQVSLKEK-----AIIQNTLBEK-----KLRVENATLKEIEELK 63
DB 591 SENKSIIEKIDVISMNASFSEVOIALNEAKLSESEKSKSCHVQOEENARLKKKKQQLQ 650
QY 64 QEL-----IOAELQNGVK-----QIAPFSGTPLHANSWSENVIGSTAVTTVSSGT 109
DB 651 QEITDWSKTLAEISEQISFEKSKQDLFEVAL--THKDDNINALTCTQTLLCESES 707
QY 110 KEQIKGGTGDEKKAKEKKEKKEKKEKQOSIAGSAD 146
DB 708 EGQNKGGNDSDDELANGEV--CGDREKKNQIKQMD 742

RESULT 15
ABB71141
ID ABB71141 standard; Protein, 1833 AA.

XX AC ABB71141;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 40215.
XX KM Drosophila; developmental biology; cell signalling; insecticide;
XX OS pharmaceutical.
XX PA Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li FWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL15244.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Disclosure; SEQ ID NO 40215; 21pp + Sequence listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
XX CC sequences (AB101840-AB16175) and the encoded proteins
XX CC (AB57737-AB572072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1833 AA;

Query Match 14.9%; Score 105.5; DB 22; Length 1833;
Best Local Similarity 27.0%; Pred. No. 1.5;
Matches 41; Conservative 24; Mismatches 56; Indels 31; Gaps 5;

QY 12 EQKGAEDQIIEYLKQVSLKEKAIQATLREBKLRVENAKLKEIEELKQELIQAET 71
DB 725 EDDPHEIRILELNEQESITRLKV-----EDLEKENAESKRYRELQAKLRQ-DS 774
QY 72 QNGVKQIAPSGTPLHAN---SMVSENVIGSTAVTTVSSGTKEQIKG-----GTGD 119

DB 775 SNGSKSILSIGTSSSAEKKVKTILNEELVQLRRTLTEKEQTVDSLKNQLSKLDLTLETEN 834
QY 120 EKAKER-----TEKKGEKKEKQOSIA 142
DB 835 DKLAKENRLLALRKASEKTEGEVDQKWKESLA 866

Search completed: January 2, 2004, 18:20:30
Job time : 45.818 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:17:12 ; Search time 11.5294 Seconds
(without alignments)
599.590 Million cell updates/sec

Title: US-09-930-169-1

Perfect score: 709
Sequence: 1 MANNDAVLKRLKRGKAEADQ.....EKKGKKKKQKQSIAGSADS 147

Scoring table: BLOSUM62
Gap 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	709	100.0	312 1 MCAL_HUMAN	Q12904 homo sapien
2	598.5	77.4	559 1 MCAL_CRICK	Q54873 cricgallus
3	509	71.8	310 1 MCAL_MOUSE	P31230 mus musculus
4	125.5	17.7	516 1 P54_EWTPC	P13692 enterococcu
5	117	16.5	662 1 TLBP_BACSU	P39217 bacillus su
6	116.5	16.4	1079 1 IR2P_SCHPO	Q10251 echizosacch
7	108.5	15.3	662 1 MCPB_BACSU	P35215 bacillus su
8	106.5	15.0	1093 1 TMF1_HUMAN	P82094 homo sapien
9	106	15.0	461 1 US45_LACIC	P22865 lactococcu
10	104	14.7	473 1 YVCE_BACSU	P40767 bacillus su
11	104	14.7	570 1 STIM_DROME	P83094 drosophila
12	104	14.7	961 1 VDP_BOVIN	P41541 bos taurus
13	103.5	14.6	561 1 EZPA_BACHD	Q94802 bacillus ha
14	103	14.5	1130 1 Y117_CAEEL	Q11102 caenorhabdi
15	103	14.5	1169 1 SMC_METUA	Q59037 mechanococ
16	103	14.5	2230 1 GOG4_HUMAN	Q13439 homo sapien
17	102	14.4	886 1 RA50_SITAC	Q33600 sulfolobus
18	102	14.4	1220 1 IR2P_HUMAN	O60841 homo sapien
19	101.5	14.3	539 1 MY53_HYDAT	P33922 hydra atten
20	101.5	14.3	966 1 H1PR_HUMAN	O75146 homo sapien
21	101	14.2	956 1 KF5C_MOUSE	P28738 mus musculu
22	101	14.2	957 1 KF5C_HUMAN	O60282 homo sapien
23	101	14.2	1938 1 MY53_CHICK	P13538 gallus galli
24	100.5	14.2	1690 1 C190_DROME	Q9485 drosophila
25	100	14.1	1243 1 SMC4_MICAR	Q96845 microtus ar
26	99.5	14.0	886 1 RA50_ARCFU	O92330 archaeoglob
27	99.5	14.0	2116 1 MY52_DICDI	P08799 dictyosteli
28	99	14.0	1940 1 MYH3_CHICK	P02565 gallus galli
29	98.5	13.9	407 1 Y173_MOUSE	O67720 aquifex aeo
30	98.5	13.9	879 1 RA50_SUTLO	Q94975 sulfolobus
31	98	13.8	919 1 INCE_HUMAN	O922m7 homo sapien
32	98	13.8	1197 1 YOIK_HUMAN	P77306 escherichia
33	97.5	13.8	553 1 YOIK_ECOLI	

34	97.5	13.8	727 1 MFPI_ARATH	Q91w85 arabidopsis
35	97	13.7	1605 1 RRB1_MOUSE	Q99p15 mus musculu
36	97	13.7	2434 1 YCF1_OENHO	Q9mth1 oenothera h
37	96.5	13.6	843 1 YMS1_YEAST	O05050 saccharomyc
38	96	13.5	856 1 CLPB_HELPY	P71404 helicobacte
39	96	13.5	1526 1 MY52_SCHPO	Q9us16 schizosacch
40	96	13.5	1534 1 RRB1_CANFA	Q28298 canis fam11
41	95.5	13.5	793 1 CALD_HUMAN	Q05682 homo sapien
42	95.5	13.5	1084 1 MY5S_RABIT	P02562 orycolagrus
43	95.5	13.5	1447 1 SG51_YEAST	P35187 saccharomyc
44	95.5	13.5	3214 1 BPA1_HUMAN	Q03001 homo sapien
45	95	13.4	579 1 G160_HUMAN	Q08378 homo sapien

ALIGNMENTS

RESULT 1
MCAL_HUMAN STANDARD; PRT; 312 AA.
AC Q12904; Q96CQ9; 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Multisynthetase complex auxiliary component p43 [Contains:
DE Inducible cyclokin monocyte activating polypeptide II (EMAP-II) (Small
DE SCFEL OR EMAP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95014290; PubMed=7929199;
RA Kao J.-I, Houck K., Fan Y., Haehnelt I., Libutti S.K., Kayton M.L.,
RA Glickschelt T., Chabot J., Nowyrod R., Greenberg S., Kuang W.J.,
RA Leung D.W., Hayward J.R., Kistel W., Heath M., Brett J., Stern D.M.,
RT "Characterization of a novel tumor-derived cytokine. Endothelial-
RT monocyte activating polypeptide II,"
RL J. Biol. Chem. 269:25106-25119(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heist F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stepien L., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein W.J., Umed T.B., Toshimiki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosa S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Faney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Boulfard G.G.,
RA Bikesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.I., Skalski S., Smallus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences,"
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RL 1- SIMILARITY: Contains 1 tRNA-binding domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and that statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>)

or send an email to license@isb-sib.ch.

CC EMBL: U0117; AAA62202.1; -.

CC EMBL: BC014051; AAH14051.1; -.

DR PDB; 1E7Z; 06-FEB-01.

DR PDB; 1EUJ; 06-SEP-00.

DR PDB; 1FLD; 07-FEB-01.

DR Genew; HGNC:10648; SCYE1.

DR MIM; 603605; -.

DR GO; GO:0005615; C:extracellular space; TAS.

DR GO; GO:0005125; F:cytokine activity; TAS.

DR GO; GO:0000049; F:RNA binding activity; TAS.

DR GO; GO:0006418; P:amino acid activation; TAS.

DR GO; GO:0006935; P:chemotaxis; TAS.

DR GO; GO:0006954; P:inflammatory response; TAS.

DR InterPro; IPRO02547; tRNA_bind.

DR Pfam; PF01588; tRNA_bind; 1.

DR PIRSF; PIRSF005381; EMAP11; 1.

DR PROSITE; PS50886; TRBD; 1.

KM Protein biosynthesis; RNA-binding; tRNA-binding; Cytokine; 3D-structure.

FT PROPRP 1 146

FT CHAIN 147 312

FT DOMAIN 151 252

FT CONFLICT 79 79 A -> P (IN REF. 2).

SQ SEQUENCE 312 AA; 34326 MW; 946310A0215F7587 CRC64;

Query Match Best Local Similarity 100.0%; Score 709; DB 1; Length 312; Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANNDVAVLRKLBKGAADQIIIEYKQVSLKKEKAILQATLRBEKKRVENAKLKEIE 60

DB 1 MANNDVAVLRKLBKGAADQIIIEYKQVSLKKEKAILQATLRBEKKRVENAKLKEIE 60

QY 61 ELKQELIQAEIQNGVQKIAFP--SGTPPLHANSMSENVIOSTAVTVSSGTEQIKGTGDE 120

DB 61 ELKQELIQAEIQNGVQKIAFP--SGTPPLHANSMSENVIOSTAVTVSSGTEQIKGTGDE 120

QY 121 KKAKEIKKEKKEKKEKKQSIAGSADS 147

DB 121 KKAKEIKKEKKEKKEKKQSIAGSADS 147

RESULT 2

MCAL_CRIGR STANDARD; PRT; 359 AA.

ID MCAL_CRIGR

AC 054873;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Multisynthetase complex auxiliary component p43 [Contains: Endothelial-monocyte activating polypeptide II (EMAP-II) (Small inducible cytokine subfamily E member 1)].

DE SCYE1.

OS Cricetus griseus (Chinese hamster).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.

OC NCBI_TaxID=10029;

OX NCBI_TaxID=10029;

RN SEQUENCE FROM N.A.

RP MEDLINE=98070438; PubMed=9405472;

RX Quevillon S., Agou F., Robinson J.-C., Miranda M.; "The p43 component of the mammalian multi-synthetase complex is likely to be the precursor of the endothelial monocyte-activating polypeptide II cytokine.";

RT J. Biol. Chem. 272:32573-32579 (1997).

CC -1- SUBUNIT: COMPONENT OF THE MULTISYNTHETASE COMPLEX WHICH IS COMPRISED OF A BIFUNCTIONAL GLUTAMYL-PROLYL-TRNA SYNTHETASE, THE MONOSPECIFIC ISOLEUCYL, LEUCYL, GLUTAMINYL, METHIONYL, LYSYL, ARGINYL, AND ASPARTYL-TRNA SYNTHETASES AS WELL AS THREE AUXILIARY

PROTEINS, P18, P48 AND P43.

-1- SIMILARITY: Contains 1 tRNA-binding domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AF021800; AAB95207.1; -.

DR InterPro; IPRO02547; tRNA_bind.

DR Pfam; PF01588; tRNA_bind; 1.

DR PIRSF; PIRSF005381; EMAP11; 1.

DR PROSITE; PS50886; TRBD; 1.

KM Protein biosynthesis; RNA-binding; tRNA-binding; Cytokine. 3D-structure.

FT DOMAIN 198 299

FT SEQUENCE 359 AA; 39601 MW; 4D868D1B65D72C23 CRC64;

Query Match Best Local Similarity 77.4%; Score 548.5; DB 1; Length 359; Matches 119; Conservative 10; Mismatches 17; Indels 3; Gaps 2;

QY 1 MANNDVAVLRKLBKGAADQIIIEYKQVSLKKEKAILQATLRBEKKRVENAKLKEIE 60

DB 47 MATNDVAVLRKLBKGAADQIIIEYKQVSLKKEKAILQATLRBEKKRVENAKLKEIE 106

QY 61 ELKQELIQAEIQNGVQKIAFP--SGTPPLHANSMSENVIOSTAVTVSSGTEQIKGTG 118

DB 107 ELKQELIQAEIQNGVQKIAFPVQSDTPVQASAVSTVISTSCIKSHSKG-G 165

QY 119 DEKKAKEIKKEKKEKKEKKQSIAGSADS 147

DB 166 DEKKAKEIKKEKKEKKEKKQSIAGSADS 194

RESULT 3

MCAL_MOUSE STANDARD; PRT; 310 AA.

ID MCAL_MOUSE

AC P12310; Q60659;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Multisynthetase complex auxiliary component p43 [Contains: Endothelial-monocyte activating polypeptide II (EMAP-II) (Small inducible cytokine subfamily E member 1)].

DE SCYE1 OR EMAP2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OC NCBI_TaxID=10090;

OX NCBI_TaxID=10090;

RN SEQUENCE FROM N.A.

RP MEDLINE=95014290; PubMed=7929199;

RX Kao J., Houck K., Fan Y., Haehnel I., Libutti S.K., Kayton M.L., Grikshchait T., Chabot J., Nowygrod R., Greenberg S., Kuang W.J., Leung D.W., Hayward J.R., Kiesel W., Heath M., Brett J., Stern D.M.; "Characterization of a novel tumor-derived cytokine. Endothelial-monocyte activating polypeptide II.";

RT J. Biol. Chem. 269:25106-25119 (1994).

RL J. Biol. Chem. 269:25106-25119 (1994).

RN SEQUENCE FROM N.A.

RP MEDLINE=22388257; PubMed=12477932;

RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D., Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F., Diatchenko L., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.D., Ustin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loughellano N.A., Peters G.J., Abramson R.D., Millhys S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

```

RA  Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA  Villalon D.K., Muzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA  Faney U., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA  Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA  Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA  Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA  Butterfield Y.S.N., Krzyzanski M.I., Skalska U., Smallus D.E.,
RA  Schnerch A., Schin J.E., Jones S.J.M., Marra M.A.,
RT  "Generation and initial analysis of more than 15,000 full-length
RT  human and mouse cDNA sequences."
RL  Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RL  [3]
RP  SEQUENCE OF 145-164 FROM N.A.
RX  MEDLINE=93015897; PubMed=1400342;
RA  Kao J., Ryan U., Brett G., Chen U., Shen H., Fan Y.-G., Godman G.,
RA  Familletti P.C., Wang F., Pan Y.-C.E., Stern D., Clausen M.;
RT  "Endothelial monocyte-activating polypeptide II. A novel
RT  tumor-derived polypeptide that activates host-response mechanisms."
RL  J. Biol. Chem. 267:20239-20247(1992).
RL  [4]
RP  CHARACTERIZATION.
RX  MEDLINE=94193665; PubMed=7545917;
RA  Kao J., Fan Y., Haehnel I., Brett J., Greenberg S., Clausen M.,
RA  Taylor M., Houck K., Kissel W., Seljeflid R., Burnier J., Stern D.;
RT  "A peptide derived from the amino terminus of endothelial-monocyte-
RT  activating polypeptide II modulates mononuclear and polymorphonuclear
RT  leukocyte functions, defines an apparently novel cellular interaction
RT  site, and induces an acute inflammatory response."
RL  J. Biol. Chem. 269:9774-9782(1994).
CC  -1- FUNCTION: ALTERS ENDOTHELIAL AND MONOCYTE FUNCTIONS, INDUCES THE
CC  MIGRATION OF MONOCYTES AND GRANULOCYTES, AND INDUCES AN
CC  INFLAMMATORY RESPONSE IN THE MOUSE FOOTPAD MODEL. EMP II ELICITS
CC  A PHLOGOGENIC RESPONSE AND, POTENTIALLY, AUGMENTS THE EFFECTS OF
CC  THE OTHER TUMOR-DERIVED CYTOKINES.
CC  -1- SUBUNIT: Monomer.
CC  -1- SIMILARITY: Contains 1 tRNA-binding domain.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC  between the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/
CC  or send an email to license@isb-sib.ch).
CC  CC
DR  EMBL; U10118; AAA62203.1; -.
DR  EMBL; BC002054; AAH02054.1; -.
DR  PIR; A55053; A55053.
DR  MGD; MGI:102774; Scytl.
DR  InterPro; IPR002547; tRNA bind.
DR  Pfam; PF01588; tRNA_bind.1.
DR  PIRSF; PIRSF005381; EMAP11.1.
DR  PROSITE; PS50886; TRBD; 1.
DR  Protein biosynthesis, tRNA-binding, tRNA-binding; Cytokine.
FT  PROPEP
FT  CHAIN
FT  FT 1 144
FT  FT 145 310
FT  FT ENDOTHELIAL-MONOCYTE ACTIVATING
FT  FT POLYPEPTIDE II.
FT  FT tRNA-BINDING.
SQ  DOMAIN 149 250
SQ  SEQUENCE 310 AA; 33997 MW; A2B8FF52A33D03A0 CRC64;
Query Match 71.8%; Score 509; DB 1; Length 310;
Best Local Similarity 76.4%; Pred. No. 1.2e-25;
Matches 113; Conservative 8; Mismatches 23; Indels 4; Gaps 3;
QY 1 MANDAVYKRLRQKGAADQIIIEYKQOVSLTKKATIIQATLREKKIRVENAKIKKEIE 60
DB 1 MANDAVYKRLRQKGAADQIIIEYKQOVSLTKKATIIQATLREKKIRVENAKIKKEIE 60
QY ELKQELIDAEIIONGVKQIAPSPGPIHLANSWMSVENVIQSTAV-ITVSSSGTKEQIKGATGD 119
DB 61 ELKQELIDAEIIONGVQYVRVHSLTPLQNCITASBSVQSPVATITASPAITAEQIK--AGE 118
QY 120 EKKAKEKIEKKGEKKKKQOQSTIASGADS 147

```

```

Db      119 EKVYKETEKKGEKKE-KQSSAASTDS 145
|||||
RESULT 4
ID      P54_ENTFC          STANDARD;          PRT;          516 AA.
AC      P13692;
DT      01-JAN-1990 (Rel. 13, Created)
DT      01-NOV-1991 (Rel. 20, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      P54 protein precursor.
OS      Enterococcus faecium (Streptococcus faecium).
OC      Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX      NCBI_TaxID=1352;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=89385998; PubMed=2780297;
RA      Fuerst P., Moesch H.-U., Solioz M.;
RL      "A protein of unusual composition from Enterococcus faecium.";
RL      Nucleic Acids Res. 17:6724-6724(1989).
CC      -1- SUBCELLULAR LOCATION: Attached to the cell wall.
CC      -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C40.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC      EMBL; X16421; CAA34442.1; ALT_INIT.
DR      PIR; S05542; S05542.
DR      MEROPS; C40_UPM; -.
DR      InterPro; IPR000064; NLPC_P60.
DR      Pfam; PF00877; NLPC_P60; 1.
KW      Signal; Cell wall.
FT      SIGNAL              1       27
FT      CHAIN               28      516
FT      ACT SITE            429      POTENTIAL.
FT      ACT SITE            516 AA; 54596 MW; 402ECMA439846D26 CRC64;
SQ      SEQUENCE 516 AA; 54596 MW; 402ECMA439846D26 CRC64;
Query Match           17.7%; Score 125.5; DB 1; Length 516;
Best Local Similarity 28.1%; Pred. No. 0.27;
Matches 41; Conservative 39; Mismatches 51; Indels 15; Gaps 5;
Oy      5 DAVLKRLFGKAEADQIIETLKQGVSLKKRKALIQATLRBCKLGRVENAKLKEIEBLQ 64
        | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db      38 DKRIADIQNODASQSGIEMLEGVSAINRKA--QDLITRKQDTLRKESADLKQEIKDLOE 95
        | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Oy      65 --ELIAQIENGVCQIAFPAGTP-----THANSMSSENIYSTAVTVTSGTGEOIKGG 116
        | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db      96 RIERRETRTIQAQERIVKYNTSYNDIVAINADSL-ADAVGRIOASTIVKANQDIWQ-- 152
        | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Oy      117 TGDEKAKAKIEKKGEKKERKQOSTA 142
        | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db      153 --QQEKDKQAVEAKKENAKQKELA 176
        | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
RESULT 5
ID      TLPB_BACSU          STANDARD;          PRT;          662 AA.
AC      P39217;
DT      01-FEB-1995 (Rel. 31, Created)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Methy-1-accepting chemotaxis protein tlpB.
GN      TLPB.
OS      Bacillus subtilis.
OC      Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX      NCBI_TaxID=1423;
RN      [1]

```

RP SEQUENCE FROM N.A.
 RC STRAIN=168 / O11085;
 RX MEDLINE=94245722; PubMed=8188684;
 RA Hanlon D.W., Ordal G.W.,
 RT "Cloning and characterization of genes encoding methyl-accepting
 RT chemotaxis proteins in *Bacillus subtilis*,"
 RL J. Biol. Chem. 269:14038-14046(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogassawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertoletti M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Bortles R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Denicof F., Devigne K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallazzi A., Galleron N.,
 RA Ghim S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holtsappel S., Hosono S., Hullio M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasaahara Y., Klaert-Blanchard M., Klein C.,
 RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Portoullik S., Prescott A.M.,
 RA Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Roche A., Roche B., Rose M., Sadie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takenchi M., Takemoshi A., Tanaka T., Terpestra P., Tognoni A.,
 RA Tosaio V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzengesser T.,
 RA Winters P., Wipat K., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zandini A.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*,"
 RT Nature 390:249-256(1997).
 RL
 CC -1- FUNCTION: CHEMOTACTIC-SIGNAL TRANSDUCERS RESPOND TO CHANGES IN THE
 CC CONCENTRATION OF ATTRACTANTS AND REPELLENTS IN THE ENVIRONMENT,
 CC TRANSDUCE A SIGNAL FROM THE INSIDE OF THE CELL, AND
 CC FACILITATE SENSORY ADAPTATION THROUGH THE VARIATION OF THE LEVEL
 CC OF METHYLATION. ALL AMINO ACIDS SERVE AS ATTRACTANTS IN
 CC B. SUBTILIS, THEY APPEAR TO CAUSE AN INCREASE IN THE TURNOVER
 CC METHYL GROUPS, LEADING TO METHYLATION OF AN UNIDENTIFIED ACCEPTOR,
 CC WHILE REPELLENTS HAVE BEEN SHOWN TO CAUSE A DECREASE IN METHYL
 CC GROUP TURNOVER. THE METHYL GROUPS ARE ADDED BY A METHYLTRANSFERASE
 CC AND REMOVED BY A METHYLESTERASE.
 CC -1- SIMILARITY: Contains 1 methyl-accepting transducer domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL nucleotide
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L29189; AAA20557.1; -
 CC EMBL; 299119; CAB15101.1; -
 CC EMBL; 299120; CAB15112.1; -
 CC PIR; D54078; D54078.
 CC HSSP; P02942; 10U7.
 CC Subtilisin; BG10862; tlpB.
 CC Interpro; IPR004010; Cache.
 CC Interpro; IPR004089; Chemtaxis_transd.
 CC Interpro; IPR003660; HAMF.
 CC Interpro; IPR003123; TarH.

DR Pfam; PF02743; Cache; 1.
 DR Pfam; PF00672; HAMF; 1.
 DR Pfam; PF00015; MCPsignal; 1.
 DR Pfam; PF02203; TarH; 1.
 DR SMART; SM00304; HAMF; 1.
 DR SMART; SM00283; MA; 1.
 DR SMART; SM00319; TarH; 1.
 DR PROSITE; PS00111; CHEMOTAXIS_TRANSDUC_2; 1.
 DR PROSITE; PS00885; HAMF; 1.
 KM Chemotaxis; Transducer; Transmembrane; Methylation; Complete proteome.
 FT DOMAIN 1 16
 FT TRANSMEM 17 37
 FT DOMAIN 38 281
 FT TRANSMEM 282 302
 FT DOMAIN 303 662
 FT DOMAIN 303 355
 FT DOMAIN 374 610
 FT MOD_RES 370 370
 FT MOD_RES 594 594
 FT MOD_RES 629 629
 FT MOD_RES 636 636
 FT MOD_RES 662 662
 SQ SEQUENCE 662 AA; 71535 MW; 91215F86293D7425 CRC64;
 Query Match 16.5%; Score 117; DB 1; Length 662;
 Best local Similarity 26.0%; Pred. No. 1.2;
 Matches 45; Conservative 26; Mismatches 64; Indels 38; Gaps 5;
 QY 5 DAVLKRLKQKADADQIIEYK--QVSLKEXKILQ-----TLREKKIR 49
 DB 463 EAVVKLEKFSKSDITSILNVINGIADQTLMLANAIEMARAGEYGRGFSVAEEVRKLA 522
 QY 50 VEMAKLKEIEELKKEIQ-----AEIQGVKQIAFPGCTPLHANSVENVYST 100
 DB 523 VQSAASAKIEBGLDIEIVAEITSLSMFQSVHVEK-----GLQITDQIAESKQLY 575
 QY 101 AVTVSSG-----TKEDIKGTDEKKAKEIKKKEKKQKQSGASAD 146
 DB 576 EMTTQISGLKQVNTVBEQSLASQGVSSAVDISVAKESAGIQDIAMAE 628
 RESULT 6
 ID PF2P SCHPO STANDARD; PRT; 1079 AA.
 AC 010251;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable translocation initiation factor Pf-2.
 GN SPAC56F6.03.
 OS Schizosaccharomyces pombe (Pisces yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 ON NCBI TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Frazer A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagele K.,
 RA James K., Jones L., Jones M., Leach S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeil C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymponrez B.,
 RA Welfjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,


```

DR Pfam; PF02743; Cache; 1.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR Pfam; PF02203; Tact; 1.
DR SMART; SM00304; HAMP; 1.
DR SMART; SM00283; MA; 1.
DR SMART; SM00319; Tact; 1.
DR PROSITE; PS00111; CHEMOTAXIS_TRANSLOC_2; 1.
DR PROSITE; PS50885; HAMP; 1.
DR Chemotaxis; Transducer; Transmembrane; Methylation; Complete proteome.
KM DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 17 37 POTENTIAL.
FT DOMAIN 38 282 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 283 303 POTENTIAL.
FT DOMAIN 304 662 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 304 356 HAMP.
FT DOMAIN 375 611 METHYL-ACCEPTING TRANSDUCER.
FT MOD_RES 371 371 DEAMIDATION AND METHYLATION (BY SIMILARITY).
FT MOD_RES 595 595 DEAMIDATION AND METHYLATION (BY SIMILARITY).
FT MOD_RES 630 630 METHYLATION (BY SIMILARITY).
FT MOD_RES 637 637 METHYLATION (BY SIMILARITY).
SQ SEQUENCE 662 AA; 71885 MW; 98DD181FB46900BC CRC64;

Query Match 15.3%; Score 108.5; DB 1; Length 662;
Best Local Similarity 26.8%; Pred. No. 4.1;
Matches 44; Conservative 27; Mismatches 56; Indels 37; Gaps 7;

QY 5 DAVLKRLKQKGAADQIIEYLK---QGVSLKRAIIQA-----TLREKKLR 49
DB 464 EAVVKGLEKSKDITSLRVINGIADQTNLALNAATEARAGESGRGFSVAEEVAKLA 523
QY 50 VENAKLKKEIEELKQELIQVIONGVQIAFPSTGPLHANSMTSENVIOSTATTYSSGT 109
DB 524 VQSDADSAKEIKELIQETV-AEID-----TSHMKREYQEV-QSGLV--VTDNT 568
QY 110 KEQIKG-----GTGDEKAKKEIKKGEKKEKQOSIASD 146
DB 569 KESFQSFMTNIEAGLQTMNSTVEQLSPRSQHVSAVSGIAD 612

RESULT 8
TMPL_HUMAN STANDARD; PRT; 1093 AA.
ID P82054;
AC 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TATA element modulatory factor (TMF).
GN TMF1.
OS Homo sapiens (human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=93028466; PubMed=1409643;
RA Garcia J.A., Ou S.-H.I., Wu F., Lusis A.J., Sparkes R.S., Gaynor R.B.;
RT "Cloning and chromosomal mapping of a human immunodeficiency virus 1
RT 'TAR' element modulatory factor.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:9372-9376(1992).
CC -!- FUNCTION: THIS PROTEIN BINDS THE HIV-1 TATA ELEMENT AND INHIBITS
CC TRANSCRIPTIONAL ACTIVATION BY THE TATA-BINDING PROTEIN (TBP).
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

CC EMBL; L01042; AAD54608.1; -.
DR PIR; A47212; A47212.
DR TRANSFAC; T00835; -.
DR Genew; HGNC:11870; TMF1.
DR MIM; 601126; -.
DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. .; TAS.
DR GO; GO:0003712; F:transcription cofactor activity; TAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; TAS.
DR GO; GO:0006366; P:transcription from Pol II promoter; TAS.
KM Transcription regulation; DNA-binding; Repressor; Coiled coil.
FT DOMAIN 439 922 COILED COIL (POTENTIAL).
FT DOMAIN 984 1092 COILED COIL (POTENTIAL).
SQ SEQUENCE 1093 AA; 123170 MW; 26133E85F4677BE CRC64;

Query Match 15.0%; Score 106.5; DB 1; Length 1093;
Best Local Similarity 22.4%; Pred. No. 9.1;
Matches 37; Conservative 33; Mismatches 50; Indels 45; Gaps 5;

QY 3 NNDVAVKRLKQKGAADQIIEYLKQGVSLKE-----KAILQATLREKKLRVENAKLK 57
DB 571 HNSNIIKKLRADKQKENVVAKLKKVKELELQHLKQVLDGKEVEKQHRENIKKLNS 630
QY 58 EIEELKQEL-----IQAEIQGVQIAFPSTGPLHANSMTSENVIOSTATTYSSGT 101
DB 631 MYERQEKDILGRQLQVMDLEEKRSIQALDSAYKEL-----TDLKNAKADSEAOEAA 685
QY 102 VITVSSGTKEQIKGIGTDEKAKKE-----IEKKGEKKEKQOSIA 142
DB 686 LSR-----EMKAKELSADLEKQAEARQOQETLA 715

RESULT 9
US45_LACIC STANDARD; PRT; 461 AA.
ID US45_LACIC
AC P22865;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Secretd 45 kDa protein precursor.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxId=1359;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-37.
RC STRAIN=M61363;
RX MEDLINE=91071599; PubMed=2123812;
RA van Asseldonk M., Rutten G., Oteman W., Sieszen R.J., de Vos W.M.,
RA Simons G.;
RT "Cloning of usp45, a gene encoding a secreted protein from
RT Lactococcus lactis subsp. lactis M61363.";
RL Gene 95:155-160(1990).
CC -!- SIMILARITY: TO E.FABCTUM P54.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

Query Match 15.0%; Score 106; DB 1; Length 461;

Best Local Similarity 27.3%, Pred. No. 4.1;
Matches 42; Conservative 31; Mismatches 71; Indels 10; Gaps 5;

QY 1 MANNDAVKRLKLEQGAEDQIIEYKQVSLKKE--ALLQATLREKGLVENEAKKE 58
DQ 33 IAKDPATISSAQSAKAKQAVDSIQSNVSLQKQKSTKQIAKIEBELALNAQIATL 92
QY 59 IEELKQELIOAEIONGVKQIAFPSTGPLHA--NSKVSENVIO--TAVTVSSGTKEQIKG 115
DQ 93 NESIKERTKLEAQARSAQVNSSATNVDVAVNSKSLTDVIOKTAITATVSANKQILE- 151
QY 116 GTGPEKAKKEIKKKEKKKKQ--STAGSADS 147
DQ 152 --QOEKQKELSQKSEYKKNVQFVSLQSIDS 183

RESULT 10
YVCE_BACSU STANDARD; PRT; 473 AA.
AC P40767; 006969;
DT 01-FEB-1995 (Rel. 31, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yvce (PSP2).
GN YVCE.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RL MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertoro M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Bouvier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codan J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaat A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Jorita B., Karamata D., Kasahara Y., Klaetr-Blanchard M., Klein C.,
RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidos A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portecelle D., Porwollik S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi Y., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemori K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsita P., Tognoni A.,
RA Tosoato V., Uchiyama S., Vandenbol M., Vannier P., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzneger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumschein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256 (1997).
[3]
RP SEQUENCE OF 1-71 FROM N.A.
RX MEDLINE=89108019; PubMed=31445906;
RA Smith H., de Jong A., Bron S., Venema G.,
RT "Characterization of signal-sequence-coding regions selected from the

RT Bacillus subtilis chromosome.";
RL Gene 70:351-361 (1988).
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY C40.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z94043; CAB08053.1; -
DR EMBL; Z99121; CAB15485.1; -
DR EMBL; M22801; AAA22817.1; ALT_FRAME.
DR PIR; F70031; F70031.
DR Subtilisin; BGI1023; YVCE.
DR InterPro; IPR000064; NLP_C_P60.
DR Pfam; PF00877; NLP_C_P60; 1.
KW Hypothetical protein; Complete proteome.
FT ACT SITE 377 377 POTENTIAL.
SQ SEQUENCE 473 AA; 5103 MM; AF544B030E683038 CRC64;
Query Match 14.7%; Score 104; DB 1; Length 473;
Best Local Similarity 22.3%; Pred. No. 5.6;
Matches 37; Conservative 42; Mismatches 57; Indels 30; Gaps 4;

QY 5 DAVKRLKLEQGAEDQIIEYKQVSLKKE-----KALLQATLREKGLRV 50
DQ 34 DEKQKIESQSEVAASSIEAKEKELTELQENQKIEKELINDKLDNLSKIEDK--E 91
QY 51 ENAKLKEIEELKQELIOAE-----IQNGVKQIAFPSTGPL-----LHANSVSENV 96
DQ 92 ENDKTKSEIKKLLKKEIKETARIEKRNELIKRVRSLQBSGSGCYIDVLGSRFGDFI 151
QY 97 IQSTAVTVSSGTKEQIKGSGTGEKAKKEIKKKEKKKKQOSIA 142
DQ 152 SRATVAVSIYDADKDLKQGEQDKAKLDESDANDLTKVEQVALA 197

RESULT 11
STIM_DROME STANDARD; PRT; 570 AA.
AC P83094; O9VX16;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Stromal interaction molecule homolog precursor.
GN STIM OR CG9126.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX Ephydroidae; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21356314; PubMed=11463338;
RA Williams R.T., Manji S.S., Parker N.J., Hancock M.,
RA Van Stekelenburg L., Eid J.-P., Senior P.V., Kazanwadel J.,
RA Shandala T., Saint R., Smith P.J., Dziadek M.A.,
RT "Identification and characterization of the STIM (stromal interaction
RT molecule) gene family: coding for a novel class of transmembrane
RT proteins.";
RL Biochem. J. 357:673-685 (2001).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benoe P.V., Berman B.P., Bhandari D., Bolhakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotcher P.,
 RA Burris K.C., Busan M.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evansgela C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howard T.J., Wei M.-H., Idegawa C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laoko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang X., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry M., Morris J., Moshire A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT The genome sequence of *Drosophila melanogaster*.
 RL Science 287:2185-2195(2000).
 RN [3]

REVIEWS.

RP STRAIN=Berkeley; PubMed=12537572;
 RC MEDLINE=22426069; PubMed=12537572;
 RX Miera S., Crosby M.A., Mungall C.J., Mathews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.E.,
 RA Smith C.D., Tuvy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celisner S.B., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RT systematic review".
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley; TISSUE=Embryo;
 RX MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guerin H., Krommiller B., Paclet J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celisner S.E.;
 RT "A *Drosophila* full-length cDNA resource".
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
 CC -1- FUNCTION: possible adhesion molecule (by similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell surface
 CC (potential).
 CC -1- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).

CC EMBL: AF328906; AAK82338.1; -
 CC EMBL: AB003500; AAF48542.2; -
 CC EMBL: AY069686; AAL39831.1; -
 CC FlyBase; FBgn0045073; Stm.

DR InterPro: IPR001660; SAM.
 DR Pfam: PF00536; SAM, 1.
 DR SMART: SM00454; SAM, 1.
 DR PROSITE: PS0105; SAM_DOMAIN, 1.
 KW Cell adhesion; Transmembrane; Coiled coil; Signal; Glycoprotein.
 FT SIGNAL 1 23
 FT CHAIN 24 570
 FT DOMAIN 24 294
 FT TRANSMEM 295 312
 FT DOMAIN 313 570
 FT DOMAIN 213 281
 FT DOMAIN 26 95
 FT DOMAIN 310 407
 FT DOMAIN 420 462
 FT CARBOHYD 212 212
 SQ SEQUENCE 570 AA; 64797 MW; B8DC791F379D0B5 CRC64;
 Query Match 14.7%; Score 104; DB 1; Length 570;
 Best Local Similarity 28.7%; Pred. No. 6.8;
 Matches 43; Conservative 23; Mismatches 64; Indels 20; Gaps 4;
 QY 1 MANNDAVLRRLQKGAADQIIIEYLK-QQVSLKKEKATLREBEKLRVENA-----K 54
 DB 324 MWDMEGLPRASLOEWKELERAMEQENVATELDRKLEKAPTSSNSDLEVOQ 383
 QY 55 LKKEIEELKOEILQAEIQNGVQIAFPSCGPLHANSVSENVIOSTAVTTVSSGTKEQIK 114
 DB 384 LKKEIEMLNELSRNFE-----LVNDCWSPPPQGLSWLYQTYELESNGHX 430
 QY 115 GGTGEKAKAKKEIKKGEKKKKQOSIAGS 144
 DB 431 KRTSAEKQL-QSAREACEKLRKRSLSVGA 459

RESULT 12

ID VDP BOVIN STANDARD; PRT; 961 AA.

AC P41541;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE General vesicular transport factor p115 (Transcytosis associated
 DE protein) (TAP) (Vesicle docking protein).
 GN VDP.

OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.

NCBI TaxID=9913;

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=Kidney.
 RX MEDLINE=95132632; PubMed=7831323;
 RA Sappstein S.K., Walter D.M., Grosvener A.R., Heuser J.E.,
 RA Waters M.G.;

RT "p115 is a general vesicular transport factor related to the yeast
 RT endoplasmic reticulum to Golgi transport factor Usolp".
 RL Proc. Natl. Acad. Sci. U.S.A. 92:522-526(1995).

RN [2]

RP CHARACTERIZATION.

RX MEDLINE=92381091; PubMed=1512287;
 RA Waters M.G., Clary D.O., Rothman J.E.;
 RT "A novel 115-kD peripheral membrane protein is required for
 RT intercompartmental transport in the Golgi stack".

RL J. Cell Biol. 118:1015-1026(1992).

CC -1- FUNCTION: GENERAL VESICULAR TRANSPORT FACTOR REQUIRED FOR
 CC INTERCOMPARTMENTAL TRANSPORT IN THE GOLGI STACK; IT IS REQUIRED FOR
 CC TRANSCYTOTIC FUSION AND/OR SUBSEQUENT BINDING OF THE VESICLES TO
 CC THE TARGET MEMBRANE. MAY WELL ACT AS A VESICULAR ANCHOR BY
 CC INTERACTING WITH THE TARGET MEMBRANE AND HOLDING THE VESICULAR
 CC AND TARGET MEMBRANES IN PROXIMITY.

CC -1- SUBUNIT: HOMODIMER WITH TWO N-TERMINAL HEADS AND A C-TERMINAL
 CC COILED-COIL TAIL. DIMER FORMED BY PARALLEL ASSOCIATION OF THE

```
CC CC
CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN WHICH RECYCLES
CC BETWEEN THE CYTOSOL AND THE GOLGI APPARATUS DURING INTERPHASE.
CC -1- DOMAIN: COMPOSED OF A GLOBULAR HEAD, AN ELONGATED TAIL (COILED-
CC COIL) AND A HIGHLY ACIDIC C-TERMINAL DOMAIN.
CC -1- PPM: PHOSPHORYLATED IN A CELL CYCLE-SPECIFIC MANNER;
CC PHOSPHORYLATED IN INTERPHASE BUT NOT IN MITOTIC CELLS.
CC PHOSPHORYLATED PROTEIN ASSOCIATES WITH THE GOLGI MEMBRANE;
CC PHOSPHORYLATION PROMOTES DISSOCIATION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14186; AAA62631.1; -
CC InterPro; IPR000225; Armadillo.
CC InterPro; IPR006955; USO1_p115_C.
CC InterPro; IPR006953; USO1_p115_head.
CC Pfam; PF04871; USO1_p115_C_1.
CC Pfam; PF04869; USO1_p115_head; 1.
CC PROSITE; PS50176; ARM_REPEAT; UNKNOWN 1.
CC Transprot; Protein transport; Golgi stack; Membrane; Coiled coil;
CC Phosphorylation.
CC KW DOMAIN 1 641 GLOBULAR HEAD.
CC FT DOMAIN 642 929 COILED COIL (POTENTIAL).
CC FT DOMAIN 934 961 ASP/GLU-RICH (ACIDIC).
CC FT MOD_RES 941 941 PHOSPHORYLATION (BY SIMILARITY).
CC SQ SEQUENCE 961 AA; 107514 MW; 993FEB790ABC0AC CRC64;

Query Match 14.7%; Score 104; DB 1; Length 961;
Best Local Similarity 23.3%; Pred. No. 11;
Matches 42; Conservative 35; Mismatches 45; Indels 58; Gaps 7;

QY 1 MANDAVKRLKQGAADQIIEYKQGVSLK-----EKALIQ--ATLBEKK---L 48
DB 645 LEQHDSTVTHKMMIREDDQLQELKQIISTLKCONEDLOTAIVDOVSQIQOHQDQVLL 704
QY 49 RV-----ENAKLKEIEBLK--QELIQAEIONGVQKQIAFPS 82
DB 705 KVGQKDSQHQGYTDGQNMNGVQPEEISRLBEIEBLKSRRELLQSLAE----- 755
QY 83 GTPHANSMSENVIGSTAVTVSSGKKEQIKGGTD-EKQAKKEIKKKEKKEKQOSI 141
DB 756 -----KDSLEIKKSSQLSPGTWQSSATGADSEQIAELKQELATLKQSQINSQSV 805

RESULT 13
EZRA_BACHD STANDARD; PRT; 561 AA.
AC Q9K802;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Separation ring formation regulator.
GN EZRA OR BH3205.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kihara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
```

```
CC CC
CC -1- FUNCTION: NEGATIVE REGULATOR OF FT2Z RING FORMATION; MODULATES THE
CC FREQUENCY AND POSITION OF FT2Z RING FORMATION. INHIBITS FT2Z RING
CC FORMATION AT POLAR SITES. INTERACTS EITHER WITH FT2Z OR WITH ONE
CC OF ITS BINDING PARTNERS TO PROMOTE DEPOLYMERIZATION (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH THE CELL MEMBRANE AND ALSO
CC COLOCALIZED WITH FT2Z TO THE NASCENT SEPTAL SITE (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP001518; BAB06924.1; -
CC PIR; E84050; E84050.
CC KW Cell division; Septation; Transmembrane; Coiled coil;
CC Complete proteome.
CC FT TRANSMEM 1 21 POTENTIAL.
CC FT DOMAIN 94 133 COILED COIL (POTENTIAL).
CC FT DOMAIN 194 214 COILED COIL (POTENTIAL).
CC FT DOMAIN 251 466 COILED COIL (POTENTIAL).
CC SQ SEQUENCE 561 AA; 65161 MW; C5C315CA960F0666 CRC64;

Query Match 14.6%; Score 103.5; DB 1; Length 561;
Best Local Similarity 26.4%; Pred. No. 7.2;
Matches 47; Conservative 23; Mismatches 53; Indels 55; Gaps 6;

QY 5 DAVLKRLKQGA-----ADQIIEYKQGVSLKKEKALIQAT 41
DB 254 DQWERLEKRVLLDQTVLECGMEEINFTSESGMPELLEKVEKNETTILLPN 313
QY 42 LRE-----ENAKLKEIEBLKQEL-IOAEIONGV 75
DB 314 LREDLTKTEELTLTKETESVQLSYRLAEELVFOQKLGKELKELIQQLVDEVTBEQ 373
QY 76 KQIAFPGTPLHNSMSENVIGSTAVTVSSGKKEQIKGGTD-EKQAKKEIKKKEKKEK 133
DB 374 KQ-TFSS-----VRSMLEWRREGITACONKIEQAQESLNSLRDELKAKELKQKKEK 425

RESULT 14
YL17_CAEEL STANDARD; PRT; 1130 AA.
AC Q11102;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 131.5 kDa protein CO2P12.7 in chromosome X.
GN CO2P12.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Miller N.;
DE Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: WEAK, TO MYOSINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U41545; AAK39135.1; -
CC PIR; T34081; T34081.
```

DR WormPep; CO2F12.7; CE03901.
 KW Hypothetical protein; Coiled coil.
 FT DOMAIN 121 779 COILED COIL (POTENTIAL).
 FT DOMAIN 805 1061 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1130 AA; 131485 MW; B0P2DFE3D99FB09 CRC64;
 Query Match 14.5%; Score 103; DB 1; Length 1130;
 Best Local Similarity 20.6%; Pred. No. 16;
 Matches 37; Conservative 44; Mismatches 59; Indels 40; Gaps 5;
 QY 4 NDAVLRLEQKGAADQIIIEYLK---QVSLKEKAILQATLREKKLRVENAKLKEK- 59
 DB 870 DNSIQEKIKERKATINEMERLKSRENELAKLHEEMYMOKT--QNEKREBESKLPQELM 927
 QY 60 -----ELKQELIQAEIQ-----NGVKQIAFPGSGTPLHANSWSEVNIQSTAV 102
 DB 928 FEKEQLEAEKAEQSHLEAEVEQVQADKESKKEQLEDLENLQKRELIIQQLQDPTDE 987
 QY 103 TTVSSGTEKQI-----KGGTDEKAKKIKKKEKKKKQOSIAGSA 145
 DB 988 STEPHTKKMSITSHGVFQNFVQMKDKKEASEKRTREAKKAKKEKKEKAKAKA 1047
 RESULT 15
 SMC METUA STANDARD; PRT; 1169 AA.
 AC 059037;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Chromosome partition protein smc homolog.
 GN M1643.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcales; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.,
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii,"
 RT Science 273:1058-1073(1996).
 RL [2]
 RN REVISIONS.
 RP Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.,
 RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 RL -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN CHROMOSOME STRUCTURE AND
 PARTITIONING. ESSENTIAL FOR CHROMOSOME PARTITION (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; U67604; AAB9663.1; -
 DR TIGR; M1643; -
 DR InterPro; IPR003439; ABC transporter.
 DR InterPro; IPR005289; GTP-binding_dom.
 DR InterPro; IPR003405; SMC_C.
 DR InterPro; IPR003395; SMC_N.
 DR Pfam; PF02483; SMC_C_1.
 DR Pfam; PF02463; SMC_N_1.
 DR TIGRPFAM; TIGR00650; MG442; 1.
 DR Hypothetical protein; ATP-binding; Coiled coil; Complete proteome.
 FT NP_BIND 31 38
 FT DOMAIN 160 521 COILED COIL (POTENTIAL).
 FT DOMAIN 673 1032 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1169 AA; 136634 MW; B63CCE34E4C03F36 CRC64;
 Query Match 14.5%; Score 103; DB 1; Length 1169;
 Best Local Similarity 24.0%; Pred. No. 16;
 Matches 35; Conservative 32; Mismatches 31; Indels 48; Gaps 6;
 QY 3 NNDVLRLEQKGAADQIIIEYLK---QVSLKEKAILQATLREKKLRVENAKLKE 58
 DB 281 NN--IINELNKGNE--EVLHKSIKELVEIENDKVLDSINELKKVEIENKKKE 336
 QY 59 IELKQELIQAEIQONGVQIAFPSSGTPHANSWSEVNIQSTAVTVSSGTEKQIKGTG 118
 DB 337 IKETQKTIENR-----DSILE-----KEQ----- 356
 QY 119 DEKAKKKEKKEKKEKKKQOSIAGS 144
 DB 357 QIKELIEKIKINLYEKERLKEAIAES 382

Search completed: January 2, 2004, 18:21:03
 Job time : 13.5294 secs

Result	Query	Score	No.	Description
ID	Length	Match		
08C2U7	319	71.2	1	08C2U7 mus musculus
09KUJ3	576	17.7	2	09KUJ3 enterococcus
0589U7	270	17.6	3	0589U7 pyrococcus
09V5O4	323	16.9	4	09V5O4 dirosophila
08ENJ2	460	16.1	5	08ENJ2 ceanoabacti
08IOV8	971	15.8	6	08IOV8 osceomodum
0145Z7	1024	15.7	7	0145Z7 homo sapien
09UDT6	1066	15.7	8	09UDT6 homo sapien
08T5C7	2055	15.7	9	08T5C7 plasmodium
08ITP3	965	15.7	10	08ITP3 plasmodium
08IIG7	964	15.7	11	08IIG7 plasmodium
09JH24	837	15.6	12	09JH24 rattus norv
P90895	947	15.5	13	P90895 ceanoabadi
09XU4	244	15.4	14	09XU4 ceanoabadi
0962I6	2385	15.4	15	0962I6 plasmodium
Q06166	1661	15.3	16	Q06166 plasmodium

17	108	15.2	456	16	Q9CJH1	Q9cjj1 lactococcus
18	108	15.2	1087	5	Q969Z3	Q969z3 dictyostella
19	107	15.1	405	2	Q9RH46	Q9rh46 coxiella bu
20	107	15.1	720	4	Q9H6Q7	Q9h6q7 homo sapien
21	107	15.1	1690	5	Q449Z9	Q449z9 dirosophila
22	106.5	15.0	1372	12	Q66275	Q66275 cassava vel
23	106.5	15.0	1372	12	Q66283	Q66283 cassava vel
24	106	15.0	204	5	Q85YK9	Q85y93 dirosophila
25	106	15.0	404	11	Q8CGD6	Q8cg64 mus musculu
26	105.5	14.9	806	11	Q8VMD4	Q8vmd4 mus musculu
27	105.5	14.9	1193	4	Q9Z580	Q9z580 homo sapien
28	105.5	14.9	1833	5	Q9VM67	Q9vm67 dirosophila
29	105.5	14.9	5476	5	Q9NJJ7	Q9nj17 dirosophila
30	105.5	14.9	5533	5	Q9VPI2	Q9vp12 dirosophila
31	105.5	14.9	5553	5	Q9U6C3	Q9u6c3 dirosophila
32	105.5	14.9	5554	5	Q9NHN1	Q9nhn1 dirosophila
33	105.5	14.9	5560	5	Q9VPI1	Q9vp11 dirosophila
34	105	14.8	450	11	Q9DA19	Q9da19 mus musculu
35	105	14.8	607	5	Q8IEM7	Q8iem7 plasmodium
36	105	14.8	1900	5	Q8IKM3	Q8ikm3 dirosophila
37	105	14.8	2028	5	Q9VCD1	Q9vcd1 dirosophila
38	104.5	14.7	612	5	Q8STJ9	Q8stj9 encephalito
39	104	14.7	516	17	Q58294	Q58294 methanococc
40	104	14.7	739	2	Q9RQJ4	Q9rg4 streptococc
41	104	14.7	1510	5	Q259Z0	Q259z0 plasmodium
42	103.5	14.6	1012	11	Q8CRU1	Q8crh1 mus musculu
43	103.5	14.6	1047	11	Q9EBR1	Q9ebr1 mus musculu
44	103	14.5	286	16	Q24944	Q24944 helicobacter
45	103	14.5	405	2	Q45835	Q45835 coxiella bu

ALIGNMENTS

RESULT

ID	ORG2UT7	PRELIMINARY;	PRT;	319 AA.
AC	ORG2UT7			
RT	01-MAR-2003	(T-EMBLrel. 23, Created)		
DT	01-MAR-2003	(T-EMBLrel. 23, Last sequence update)		
DT	01-MAR-2003	(T-EMBLrel. 23, Last annotation update)		
DE	Endothelial monocyte activating polypeptide 2.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ND; TISSUE=Thymus;			
RC	MEDLINE=22354683; PubMed=12466651;			
RA	The FANTOM Consortium,			
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;			
RT	"Analysis of the mouse transcriptome based on functional annotation of			
RT	60,770 full-length cDNAs."			
RL	Nature 420:563-573(2002);			
RL	EMBL; AK087932; BAC04045.1;			
SO	SEQUENCE	319 AA; 35197 bp;	B977B8FE742BEBC	CRC64;

	Query Match	71.2%;	Score 505;	DB 11;	Length 319;
	Best Local Similarity	75.7%;	Pred. No. 1.2e-26;		
	Matches 112;	Conservative	9;	Mismatches 23;	Indels 4; Gaps 3
QY	1	MANDAVLKLRLKQGAADQIIEYLKKCVSLLEKAKIIQATLRREKULRVENAKLKELE	60		
Dd	10	MATNDVAVLKRLQGKAADQIIEYLKKQVALLLKEKAIQATLMREKULRVENAKLKEILE	69		
QY	61	EIKQELIOAELQNGYKOIAFSPSGIPHLANSWSENVIOSTLV-TTVSSGTEIQIKGGTGD	119		
Dd	70	EIKQELILAEIHNGVEQVRVLSTPLQTNCRASBSVQSPVAATTAATLTKEQIR--AGE	127		
QY	120	EKKAKEIKERKGGEKKKKQOOSIASADS	147		
Dd	128	EKKVKEKERKGGEKKE-KQSSAAASTDS	154		

RESULT 2
ID Q9KJ3 PRELIMINARY; PRT; 576 AA.
AC Q9KJ3;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Secreted antigen SagB.
GN SAGB.
OS Enterococcus hirae.
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
ON NCBI_TaxID=1354;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IX2751;
RA Teng F., Murray B.E., Weinstock G.M.;
RT "Characterization of an enterococcal secreted antigen."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF167576; AAF87093.1; -
DR InterPro: IPR000064; NLP_C_P60.
DR Pfam: PF00877; NLP_C_P60; 1.
SQ SEQUENCE 576 AA; 60807 MW; 76F453E2780B57A7 CRC64;
Query Match 17.7%; Score 125.5; DB 2; Length 576;
Best Local Similarity 28.1%; Pred. No. 0.81; Matches 41; Conservative 39; Mismatches 51; Indels 15; Gaps 5;
QY 5 DAVLKRLQKGAENDQIIEYLKQVSLKKEKALQATLREKKLRVENAKLKEIEELKQ 64
DB 38 DKXIDLQNGQASQSGIEALLEGVSAINTKA--QDLTKQDTLRKESADLKEIDLQ 95
QY 65 --ELIQAEIQNGVKQIAFPBGT-----LHANSMSSENVITQSTAVTSSGTRKQIKG 116
DB 96 RIKRREKTLIKQKARETVKNTSSNYIDAVIANADSL-ADAVRIQAMSTIVKANQDILVQ-- 152
QY 117 TGDEKKAKEIKKEKKEKKKQOOSIA 142
DB 153 --QKEDKQAVEAKKAKENAKOKELA 176
RESULT 3
ID 058907 PRELIMINARY; PRT; 270 AA.
AC 058907;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein PH1189.
GN PH1189.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
ON NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RA MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosegi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuho Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeabacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
KM EMBL: AP000005; BAA30289.1; -
SQ SEQUENCE 270 AA; 31019 MW; AA54AC8B94FBIEFE CRC64;
Query Match 17.6%; Score 124.5; DB 17; Length 270;

Best Local Similarity 25.4%; Pred. No. 0.43;
Matches 44; Conservative 34; Mismatches 44; Indels 51; Gaps 6;
QY 3 NDAVKRLQK-----GAEDQIIEYLKQVSLKKEKALQATLREKKLRVENAKL 55
DB 86 NNVAIISSEKPEFVKFSENESTQKLTKEENELKQ--IQNTKEEDLKQENADL 143
QY 56 KKEIEELKQELIQAEIQNGVKQIAFPBGTPLHANSMSSENVITQSTAVTSSGTRKQIKG 115
DB 144 KRISDLKRLKAKAQD-----ISLQVQIINLTENNELKELKAN 185
QY 116 GTG--DEKAKK-----IEKKKKKK-----KQSGIAGS 144
DB 186 QTNITNLAKAKKFLKQNNREYRLITKLREQAKKSEGSYIEKAKREKLIGS 238
RESULT 4
ID Q9V504 PRELIMINARY; PRT; 323 AA.
AC Q9V504;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE CG8235 protein.
GN CG8235.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
ON Ephydriidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Morten J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Milos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bencos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.O., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Modarres C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J.R., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong K., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Weissman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [2]


```

RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Batzon J., An H., Baldwin D., Banson J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorett V., Doup L.E., Doyle C., Dreene D., Farrant D.,
RA Ferreira S., Frise E., Galle R.F., Gary N.S., George R.A.,
RA Gonzalez M., Houch J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Matrei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J.,
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B., Rector F.,
RA Phuanavong S., Pitterman G.S., Puri V., Richards S., Scheeler F.,
RA Stepleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Smu S., Smutnick F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB003835; AAF59019.2; -
DR FlyBase; FBgn0033351; CG8235.
DR InterPro; IPR002547; tRNA_bind.
DR Pfam; PF01588; tRNA_bind; 1.
SQ SEQUENCE 323 AA; 34401 MW; 6EBB6E30115F4E84 CRC64;

Query Match 16.9%; Score 119.5; DB 5; Length 323;
Best Local Similarity 24.8%; Pred. No. 1.1;
Matches 37; Conservative 33; Mismatches 52; Indels 27; Gaps 4;

QY 6 AVLRLBEGKGAADQITLYLKQVSLKEXALIQATLRBEKKLRVENAKLKEIEELKOE 65
DB 31 ADLQOIALNNRERAEALINSIAEISGIGQ---QLVERQKQELIKENALKEVEALALQ 86
QY 66 LIOAEIQNGVQKQIAFP-----SGTPLHANSVSENVIOSTAVTTVSSGTKEQIKGGTG 118
DB 87 LVQLELRNGKQKQIIVPGARGFCTSAAPV---VWPAEAGPATAPAAFA----- 131
QY 119 DEKRAKEKIEKKERKKQKQSIAGSADS 147
DB 132 -PKPAKEKPEKKEKKEKPAEKPAAPEA 159

RESULT 5
Q8ENJ2 PRELIMINARY; PRT; 460 AA.
AC Q8ENJ2;
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Hypothetical conserved protein.
GN OR2491.
OS Oceanobacillus theyensis.
OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.
CX NCBI_TaxID=162710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;

```

```

RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus theyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments."
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AP004601; BAC14447.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 460 AA; 51501 MW; 28D97339F99CA66 CRC64;

Query Match 16.1%; Score 114; DB 16; Length 460;
Best Local Similarity 29.4%; Pred. No. 3.8;
Matches 45; Conservative 28; Mismatches 54; Indels 26; Gaps 7;

QY 9 KRLBQKGAENDQI---IEYLKQVSLKEXALIQATLRBEKKLRVENAKLKEIEELKOE 65
DB 90 KQLETKQTEIDNTNDENSLTEKVEELKER--MKLEIEELKIKE--RIEKDELKLNK 144
QY 66 LIOAEIQNGVQKQIAFP-----SGTPLHANSVSENVIOSTAVTTVSSGTKEQIKGGTG 121
DB 145 LLSIQ-QNGGDIKYLEVIFGA-----SSFGDLSRSTAVTVMQDKEIMEGERDKQL 198
QY 122 -----KAKKEKKEKKEKKEKQSIAGSAD 146
DB 199 EENKEVVASKEVEEKKLTVEEKKSLBEGHD 231

RESULT 6
Q8IOV8 PRELIMINARY; PRT; 971 AA.
AC Q8IOV8;
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE ATP-depct. acyl-CoA synthetase; putative (EC 6.2.1.3).
GN PF0085C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
CX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Cotton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray A., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabbittowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
RL Nature 419:527-531(2002).
DR EMBL; AL034557; CAD949109.1; -
KW Ligase.
SQ SEQUENCE 971 AA; 110796 MW; 867BE53D8F4B6B25 CRC64;

Query Match 15.8%; Score 112; DB 5; Length 971;
Best Local Similarity 25.8%; Pred. No. 11;
Matches 39; Conservative 32; Mismatches 44; Indels 36; Gaps 6;

QY 5 DAVKRLBEGKGAENDQI-----QIEYLKQVSLKEXALIQATLRBEKKLRVE 51
DB 787 DQVKKTEEDKLGKSGSTSSKSVSGGKKERKENDIGENKVASVEKNEKVEKSEVK 846
QY 52 NAKLKEIE-ELKQEL---IOAEIQNGVQKQIAFPSSGTPLHANSVSENVIOSTAVTTVSS 107

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:18:32 ; Search time 16.112 Seconds
(without alignments)
875.822 Million cell updates/sec

Title: US-09-930-169-1

Perfect score: 1 MANNDAVLKRLBQKGAADQ.....EKKGKKKKQOSIAGSADS 147

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	682	96.2	310	2 B55053	endothelial monocy
2	509	71.8	310	2 B55053	endothelial monocy
3	125.5	17.7	507	2 S05342	hypothetical prote
4	124.5	17.6	270	2 G71061	hypothetical prote
5	117	16.5	662	2 D54078	methy-accepting c
6	116.5	16.4	1079	2 T38913	translational initia
7	110	15.5	947	2 T20156	hypothetical prote
8	109.5	15.4	244	2 T26265	hypothetical prote
9	108.5	15.3	662	2 A54078	methy-accepting c
10	108	15.2	456	2 B86903	hypothetical prote
11	108	15.2	461	2 JN0097	secreted 45k prote
12	108	15.2	1087	2 T30330	gelsoilin-related p
13	107	15.1	1690	2 T13030	microtubule bindin
14	106.5	15.0	1093	2 A47212	transcription fact
15	104	14.7	473	2 F70031	cell wall-binding
16	104	14.7	516	2 D64410	replication factor
17	104	14.7	1526	2 A45605	mature-parasite-in
18	103.5	14.6	561	2 B84050	hypothetical prote
19	103	14.5	286	2 B64536	hypothetical prote
20	103	14.5	1130	2 T34081	hypothetical prote
21	103	14.5	1169	2 A64505	hypothetical prote
22	102.5	14.5	853	2 T51505	PL15 homolog - Met
23	102	14.4	405	2 S42875	hypothetical prote
24	102	14.4	530	2 B82189	diacylglycerol S
25	102	14.4	1132	2 T34383	translational initia
26	101.5	14.3	433	2 T71853	hypothetical prote
27	101.5	14.3	1046	2 T42720	cytoplasmic linker
28	101	14.2	1365	2 T45031	hypothetical prote
29	101	14.2	1938	1 JX0178	myosin heavy chain

30	100	14.1	777	2 T21048	hypothetical prote
31	99.5	14.0	886	2 H69378	conserved hypothet
32	99.5	14.0	2116	2 A26655	myosin heavy chain
33	99	14.0	392	2 G95258	secreted 45 kd pro
34	99	14.0	392	2 B98124	general stress pro
35	99	14.0	1388	2 T30335	KUP2 protein - Afr
36	99	14.0	1940	2 A29320	myosin heavy chain
37	98.5	13.9	407	2 F70461	hypothetical prote
38	98.5	13.9	1046	2 T42734	cytoplasmic linker
39	98.5	13.9	1085	2 F96712	hypothetical prote
40	98.5	13.9	1115	2 B84476	probable TPR repea
41	98.5	13.9	1179	2 F71190	probable chromosom
42	98.5	13.9	1827	2 T16270	hypothetical prote
43	98	13.8	1269	2 F84730	probable myosin he
44	97.5	13.8	553	2 E91120	probable membrane
45	97.5	13.8	553	2 E85965	probable membrane

ALIGNMENTS

```
RESULT 1
B55053
endothelial monocyte-activating protein II precursor - human
C/Species: Homo sapiens (man)
C/Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 09-Dec-2002
C/Accession: B55053
R/Kao, J.; Houck, K.; Fan, Y.; Haehnel, I.; Libutti, S.K.; Kayton, M.L.; Grikscheit, T.;
J.; Stern, D.M.
J. Biol. Chem. 269, 25106-25119, 1994
A/Title: Characterization of a novel tumor-derived cytokine. Endothelial-monocyte activa
A/Reference number: A55053; MUID:95014290; PMID:7929199
A/Accession: B55053
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-310 <KAO>
A/Cross-references: GB:U10117
C/Superfamily: endothelial monocyte-activating polypeptide II precursor (pro-EMAP II)

Query Match 96.2%; Score 682; DB 2; Length 310;
Best Local Similarity 98.0%; Pred. No. 1.5e-37;
Matches 144; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

Cy 1 MANNDAVLKRLBQKGAADQIEYLKQOVSLKEKATLQATLREKKLRVENAKLKEIE 60
Db 1 MANNDAVLKRLBQKGAADQIEYLKQOVSLKEKATLQATLREKKLRVENAKLKEIE 60

Cy 61 ELKQELIQAEIQNGVQKQIRFPGSGPLHANSMSENVISTAVTTVSSGTGEQIKGTGDE 120
Db 61 ELKQELIQAEIQNGVQKQIRFPGSGPLHANSMSENVISTAVTTVSSGTGEQIKG--GDE 118

Cy 121 KKAKETIEKKGEKKKKQOSIAGSADS 147
Db 119 KKAKETIEKKGEKKKKQOSIAGSADS 145

RESULT 2
A55053
endothelial monocyte-activating protein II precursor - mouse
C/Species: Mus musculus (house mouse)
C/Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 09-Dec-2002
C/Accession: A55053; A44032
R/Kao, J.; Houck, K.; Fan, Y.; Haehnel, I.; Libutti, S.K.; Kayton, M.L.; Grikscheit, T.;
J.; Stern, D.M.
J. Biol. Chem. 269, 25106-25119, 1994
A/Title: Characterization of a novel tumor-derived cytokine. Endothelial-monocyte activa
A/Reference number: A55053; MUID:95014290; PMID:7929199
A/Accession: A55053
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-310 <KAO>
A/Cross-references: GB:U10118; NID:G498911; PIDN:AAA62203.1; PID:G498912
R/Kao, J.; Ryan, J.; Brett, G.; Chen, J.; Shen, H.; Fan, Y.G.; Godman, G.; Familletti, P
```

J. Biol. Chem. 267, 20239-20247, 1992
A>Title: Endothelial monocyte-activating polypeptide II. A novel tumor-derived polypeptide
A/Reference number: A44032; MUID:93015897; PMID:1400342
A/Accession: A44032
A/Molecule type: Protein
A/Residues: 145-158, 'X', 160-164 <KA2>
A/Experimental source: methylcholanthrene A fibrosarcoma cells
A/Note: sequence extracted from NCBI backbone (NCBI:115676)
C/Superfamily: endothelial monocyte-activating polypeptide II precursor (pro-EMAP II)

Query Match 71.8%; Score 509; DB 2; Length 310;
Best Local Similarity 76.4%; Pred. No. 2, 7e-26;
Matches 113; Conservative 8; Mismatches 23; Indels 4; Gaps 3;

QY 1 MANDAVLKRLKLEOKGAEADQIIIEYLKQVSLKKEKAILQATLREKKLRVENAKLKEIE 60
DB 1 MANDAVLKRLKLEOKGAEADQIIIEYLKQVSLKKEKAILQATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIQAELIONGVQKQIAPPSGTPPLHANSWSENVIOSTAV-TTVSSGTKEQIKGCTGD 119
DB 61 ELKQELIQAELIONGVQKQIAPPSGTPPLHANSWSENVIOSTAV-TTVSSGTKEQIKGCTGD 119
QY 120 EKRAKEIKKKEKKEKKQKQSIAGSADS 147
DB 119 EKRAKEIKKKEKKEKKQKQSIAGSADS 145

RESULT 3
505542
hypothetical protein, 54K - Enterococcus faecium
C/Species: Enterococcus faecium
C/Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 15-Oct-1999
C/Accession: S05542
R/Fuerset, P.; Moesch, H.U.; Solter, M.
Nucleic Acids Res. 17, 6724, 1989
A>Title: A protein of unusual composition from Enterococcus faecium.
A/Reference number: S05542; MUID:89385998; PMID:2780297
A/Accession: S05542
A/Molecule type: DNA
A/Residues: 1507 <FUE>
A/Cross-references: GB:X16421; EMBL:M26048; NID:g43333; PIDN:CAA34442.1; PID:g43334
A/Note: the authors translated the codon CGT for residues 221 and 223 as Lys

Query Match 17.7%; Score 125.5; DB 2; Length 507;
Best Local Similarity 28.1%; Pred. No. 0.38;
Matches 41; Conservative 33; Mismatches 51; Indels 15; Gaps 5;

QY 5 DAVLKRLKLEOKGAEADQIIIEYLKQVSLKKEKAILQATLREKKLRVENAKLKEIEELKO 64
DB 29 DKTIADLQNGQASQISQIEALLEGVSAINTKA--QDLITKQDTLRKESAGLKQEIKDLOE 86
QY 65 --ELIQAEIIONGVQKQIAPPSGTP-----LHANSWSENVIOSTAVTTVSSGTKEQIKG 116
DB 87 RIKRREATTIKQARETQVKNSTSNVYDAVINADSL-ADAVGRIOASTIVKANQDILVQ-- 143
QY 117 TGDEKAKKEIKKKEKKKQKQSI 142
DB 144 --QKEDKQAVKAKENAKQKELA 167

RESULT 4
G71061
hypothetical protein PH189 - Pyrococcus horikoshii
C/Species: Pyrococcus horikoshii
C/Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999
C/Accession: G71061
R/Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Seki,
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
DNA Res. 5, 55-76, 1998
A>Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A/Reference number: A71000; MUID:98344137; PMID:9679194
A/Accession: G71061
A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA
A/Residues: 1-270 <KAW>
A/Cross-references: GB:AP000005; NID:g3236132; PIDN:BA30289.1; PID:d1031232; PID:g32576
A/Experimental source: strain OT3
A/Note: this accession replaces an interim accession for a sequence replaced by GenBank
C/Genetics:
A/Genes: PH189

Query Match 17.6%; Score 124.5; DB 2; Length 270;
Best Local Similarity 25.4%; Pred. No. 0.23;
Matches 44; Conservative 34; Mismatches 44; Indels 51; Gaps 6;

QY 3 MNDVAVLKRLKLEOK-----GAEADQIIIEYLKQVSLKKEKAILQATLREKKLRVENAKL 55
DB 86 NNVALISISEKRPFEVKEFSESTQSLKTLKENEKELKQ--IQNLTKENQGLKNNMAQL 143
QY 56 KKEIEELKQELIQAELIONGVQKQIAPPSGTPPLHANSWSENVIOSTAVTTVSSGTKEQIKG 115
DB 144 KRISDLEKHLKKAQD-----ISELQVQINNLTKENNELKELIAN 185
QY 116 GTG--DEKAKKE-----IEKKGEKKK-----KQOSIAGS 144
DB 186 QNTTITQLKAKAKFLKQNNERTLITLLEQAKKSEOSTYIEKAKREKLIGS 238

RESULT 5
D54078
methyl-accepting chemotaxis protein tlpB - Bacillus subtilis
N/Alternate names: Chemotaxis transducer homolog TLPB
C/Species: Bacillus subtilis
C/Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Jun-2000
C/Accession: D54078; A69724
R/Hanlon, D.W.; Ordal, G.W.
J. Biol. Chem. 269, 14038-14046, 1994
A>Title: Cloning and characterization of genes encoding methyl-accepting chemotaxis prote
A/Reference number: A54078; MUID:94245722; PMID:8188684
A/Accession: D54078
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-662 <HAN>
A/Cross-references: GB:L29189; NID:g459687; PIDN:AAA20557.1; PID:g459691
A/Note: authors translated the codon CAG for residue 10 as Trp
R/Kunst, F.; Ogasawara, N.; Moser, I.; Albertin, A.M.; Alloni, G.; Azevedo, V.; Bertier,
C.; Bron, S.; Brouillet, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi,
A.; Enllich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrari, E.
Nature 390, 249-256, 1997
A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallert,
leeh, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.;
Koester, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauel,
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,
A/Authors: Schleich, S.; Schreier, R.; Scofield, F.; Sekiguchi, J.; Sekowaka, A.; Setor,
akuchii, M.; Tanakoshi, A.; Tanaka, T.; Tepstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Whitters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A/Authors: Yoshikawa, H.F.; Zumsattel, B.; Yoshikawa, H.; Danchin, A.
A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A/Reference number: A69580; MUID:98044033; PMID:9384377
A/Accession: A69724
A/Molecule type: DNA
A/Status: nucleic acid sequence not shown; translation not shown
A/Residues: 1-662 <KUN>
A/Cross-references: GB:Z99119; GB:Z99120; GB:AL009126; NID:g2655613; PIDN:CAB15112.1; PII
A/Experimental source: strain 168
C/Genetics:
A/Genes: tlpB
C/Superfamily: probable methyl-accepting chemotaxis transducer
C/Keywords: transmembrane protein

Query Match 16.5%; Score 117; DB 2; Length 662;
Best Local Similarity 26.0%; Pred. No. 1.8;
Matches 45; Conservative 26; Mismatches 64; Indels 38; Gaps 5;

OY 5 DAVKRLKLEKGAADQIIEYK---QQVSLKEKAILQA-----TLREKKLR 49
 Db 463 EAVVKGLETKSQDITSLINVLINGIADQTNLALNAITEAARAGEYGRGFSVAEVRKLA 522
 OY 50 VENAKLKEIEELKQELIQ-----AEIQNGKQIAPFSGTPIHANSVSENVIOST 100
 Db 523 VQASDAKKEIGLQIEIVREISTSLMFQSVNHEVKE-----GLQITDQTAFSEKQIY 575
 OY 101 AVTTVSSG-----TKEQIKGTGDEKKAKIEKKEKKEKQOOSIAGSAD 146
 Db 576 EMTTQISELQINMANTVQQLSAGSQEVSASVEDISAVAKESAGQIDIAAE 628

RESULT 6

translation initiation factor if-2 - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
 C:Accession: T38913
 R:Pearson, D.; Churche, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
 submitted to the EMBL Data Library, February 1996
 A:Reference number: Z21817
 A:Accession: T38913
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1079 <PEA>
 A:Cross-references: EMBL:Z69728; PIDN:CAA93574.1; GSPDB:GN00066; SPDB:SPAC56F8.03
 C:Experimental source: strain 972h-, cosmid c56F8
 C:Genetics:
 A:Gene: SPDB:SPAC56F8.03
 A:Map position: 1
 C:Superfamily: translation elongation factor Tu homology
 F:485-612/Domain: translation elongation factor Tu homology <ETU>

Query Match 16.4%; Score 116.5; DB 2; Length 1079;
 Best Local Similarity 21.6%; Pred. No. 3.3;
 Matches 38; Conservative 41; Mismatches 59; Indels 39; Gaps 5;
 OY 3 NNDVAVLKLEKGAADQIIEYKQVSLKEKAIL-----QATLREKKLRVNA 53
 Db 246 NVTALQKMLEKRRAREEQRIRREARIAEEKRLAVEARKKEALKKKEKKEKKE 305
 OY 54 KLK-----KEIEELKQELIQAEIQNGKQIAPFSG-----TPIHAN----- 89
 Db 306 EMKQKGYLKKQKQQLAQRLOQMLESQVAVGLSNGEKKQKPYVTNNKKSRSSTS 365
 OY 90 SMVSENVIOSTAVTTVSSGTKEQIKGTGDEKKAKIEKKEKKEKQOOSIAGSA 145
 Db 366 SISSGILSSPATSISVDEPQK-----DSKDSSEKVEKETEVERKEENEAEAA 415

RESULT 7

hypochemical protein F57F5.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T20156; T22857
 R:Holt, R.
 submitted to the EMBL Data Library, August 1996
 A:Reference number: Z19230
 A:Accession: T20156
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-947 <WII>
 A:Cross-references: EMBL:Z78012; PIDN:CAB01418.1; GSPDB:GN00023; CESP:F57F5.2
 A:Experimental source: clone C52E4
 R:Harris, B.
 submitted to the EMBL Data Library, July 1996
 A:Reference number: Z19627
 A:Accession: T22857
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-947 <W12>

A:Cross-references: EMBL:Z75953; PIDN:CAB00103.1; GSPDB:GN00023; CESP:F57F5.2
 A:Experimental source: clone F57F5
 C:Genetics:
 A:Gene: CESP:F57F5.2
 A:Map position: 5
 A:Introns: 24/3; 192/2; 313/1; 466/2; 574/3; 665/2; 705/2; 738/3; 832/3; 876/3; 906/3

Query Match 15.5%; Score 110; DB 2; Length 947;
 Best Local Similarity 21.6%; Pred. No. 7.6;
 Matches 37; Conservative 43; Mismatches 59; Indels 32; Gaps 6;

OY 1 MANDVAVLKLE---QKGAADQIIEYKQVSLKEKAILQATLREKKLRVNAK 57
 Db 741 LSSSDLAIVSVETSKDSGETPRPTSSSELKVRIRREALAQEK-EEERTTKEENQKIEE 799
 OY 58 EIEELKQELIQAEIQNGKQIAPFSGTPIHANSVSENVIO 98
 Db 800 VGEHDVSEATSLDSEVSHDNNISFQMPDSIPHEDRTSLPATSEIGDAISKKLE 859
 OY 99 ----STAVTTVSSGTKEQIKGTGDEKKAKIEKKEKKEKQOOSIAGSA 145
 Db 860 KEDNSSMSSLDERTTVSAKPTT-----TRLNQKLEKKEKRSWAGSS 905

RESULT 8

hypochemical protein W07G1.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T26265
 R:Percy, C.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z20186
 A:Accession: T26265
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-244 <W1>
 A:Cross-references: EMBL:Z82076; NID:el247330; PIDN:CAB04933.1; GSPDB:GN00020; CESP:W07G
 A:Experimental source: clone W07G1
 C:Genetics:
 A:Gene: CESP:W07G1.1
 A:Map position: 2
 A:Introns: 49/3; 161/1; 210/3

Query Match 15.4%; Score 109.5; DB 2; Length 244;
 Best Local Similarity 24.8%; Pred. No. 1.9;
 Matches 37; Conservative 35; Mismatches 54; Indels 23; Gaps 5;

OY 10 RLECKGAADQIIEYKQVSL-LKEKAILQATLREKKLRVENAKLKEIEE 61
 Db 75 KKKKKKKRSGQTVQPRTEHIELGDEISIPLNEPVLEKKKKKKLVQKNLDLHGSIT 134
 OY 62 LKQEL---IQAEIQNGKQIAPF---SGTPIHANSVSENVIOSTAVTTVSSGTKEQIKG 115
 Db 135 LAETAEIETLSEIDSEDEHEHRRSATPDHRRPATKSHVAPSRASSR-----RG 185
 OY 116 GTGDEKKAKIEKKEKKEKQOOSIAGS 144
 Db 186 SRDEERREDEERKKDKKRSVSS 214

RESULT 9

AS4078
 methyl-accepting chemotaxis protein mcpB - Bacillus subtilis
 C:Species: Bacillus subtilis
 C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Jun-2000
 C:Accession: A54078; H69655
 R:Hanlon, D.W.; Ordal, G.W.
 J. Biol. Chem. 269, 14038-14046, 1994
 A:Title: Cloning and characterization of genes encoding methyl-accepting chemotaxis proteins
 A:Reference number: A54078; MUID:94245722; PMID:8188684
 A:Accession: A54078
 A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-662 <HAN>
A:Cross-references: GB:129189; NID:9459687
A:Note: the sequence in GenBank entry BACMCP1P, release 106.0 (PID:9459688) has 360-N
R:Kuner, P.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bueschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emmerison, P.T.; Eutian, K.D.; Errington, J.; Fabel, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galluzzi, A.; Gall
ieb, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsteppel, S.; Hosono, S.; Hullo, M.P.
Koetter, P.; Koningsreiter, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidas, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maesuda, S.; Maue
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portebell
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,
A:Authors: Schleicher, S.; Schuster, R.; Scifo, P.; Sedgwick, J.; Sekowski, A.; Serd
akeuchi, M.; Tamkoehi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipit, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zimstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: H69655
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-359, 'N', 361-449, 'G', 451-662 <RUN>
A:Cross-references: GB:299119; GB:299120; GB:AL009126; NID:92635613; PID:NCAB15115.1; PI
A:Experimental source: strain 168
C:Comment: Strains in which this protein is inactivated are defective in chemotaxis tow
C:Genetics:
A:Gene: mcpB
C:Superfamily: probable methyl-accepting chemotaxis transducer
C:Keywords: chemotaxis; signal transduction; transmembrane protein

Query Match 15.3%; Score 108.5; DB 2; Length 662;
Best Local Similarity 26.8%; Pred. No. 6.5; 56; Indels 37; Gaps 7;
Matches 44; Conservative 27; Mismatches 56; Indels 37; Gaps 7;

QY 5 DAVLKRLKQKGAADQIIEYLK--QGVSLKEKAILQA-----TLREKKLR 49
DB 464 EAVVKGLEKSKQITSLRVINGIADQTNLALNAEAPAGSGRGFSVAEVRKLA 533
QY 50 VENAALKEIEBELKQELIQAEIONGVKQIAFPSTGPLHANSWSENVIQSTAVTTSSGRT 109
DB 524 VQSDASAKIEIKLQIETV-AEID-----TSLHMFKEVQEV-QSGLV--VTDMT 568

QY 110 KEQIKG-----GTDEKKAKEIEKKEKKEKQKQSIAGSD 146
DB 569 KESFQSIPTMTNEIAGKIQTMNSTVEQLSPRSQVSAVSGIAD 612

RESULT 10
E86903
hypothetical protein usp45 [imported] - *Lactococcus lactis* subsp. *lactis* (strain IL1403)
C/Species: *Lactococcus lactis* subsp. *lactis*
C/Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C/Accession: E86903
R:Bojorin, A.; Mincker, P.; Manger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* s
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Molecule type: DNA
A:Residues: 1-456 <STO>
A:Cross-references: GB:AE005176; PID:912725296; PIN:AAK6327.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: usp45

Query Match 15.2%; Score 108; DB 2; Length 456;
Best Local Similarity 26.6%; Pred. No. 4.7;
Matches 42; Conservative 33; Mismatches 65; Indels 18; Gaps 6;

QY 1 MANNDVLRLEKQKGAADQIIEYLKQVSLKEKAILQATLR-----EEKLRLVEN 52

DB 33 IAKDATTSSAQAQAQVDSLSQKVSQKQ---QASTKAQIAKIESAKKLNAQI 89
QY 53 AKLKEIEBELKQELIQAEIONGVKQIAFPSTGPLHANSWSENVIQSTAVTTSSGRT 111
DB 90 ATLNEISIAE-RKTLLEAQRSAQVNSATNVMDAVNSSLTDVIOKTAITAVTSANKQ 148

QY 112 QIKGTGDEKKAKEIEKKEKKEKQKQ--SIAGSADS 147
DB 149 ME---QKEKEKELSQSEYTKVKNVQVSLQSIDS 183

RESULT 11
JN0097
secreted 45K protein precursor - *Lactococcus lactis*
C/Species: *Lactococcus lactis*
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C/Accession: JN0097
R:van Aseeldonk, M.; Rutten, G.; Oteman, M.; Slezzen, R.J.; de Vos, W.M.; Simons, G.
Gene 95, 155-160, 1990
A:Title: Cloning of usp45, a gene encoding a secreted protein from *Lactococcus lactis* aut
A:Reference number: JN0097; MUID:91071599; PMID:2123812
A:Accession: JN0097
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-461 <VAN>
A:Cross-references: GB:M35374

Query Match 15.2%; Score 108; DB 2; Length 461;
Best Local Similarity 26.5%; Pred. No. 4.8;
Matches 41; Conservative 33; Mismatches 69; Indels 12; Gaps 5;

QY 1 MANNDVLRLEKQKGAADQIIEYLKQVSLKE-----KAILQATLRREKKLRVENAKI 55
DB 33 IAKDATTSSAQAQAQVDSLSQKVSQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 92
QY 56 KKEIEBELKQELIQAEIONGVKQIAFPSTGPLHANSWSENVIQSTAVTTSSGRT 114
DB 93 NSIKR-RKTLLEAQRSAQVNSATNVMDAVNSSLTDVIOKTAITAVTSANKQ 151

QY 115 GGTDEKKAKEIEKKEKKEKQKQ--SIAGSADS 147
DB 152 ---QKEKEKELSQSEYTKVKNVQVSLQSIDS 183

RESULT 12
T30330
gelsoilin-related protein GRP125 - slime mold (*Dictyostelium discoideum*)
C/Species: *Dictyostelium discoideum*
C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C/Accession: T30330
R:Stocker, S.; Hiery, M.; Marriot, G.
Mol. Biol. Cell 10, 161-178, 1999
A:Title: Phototactic migration of *Dictyostelium* cells is linked to a new type of gelsoilin
A:Reference number: Z20823; MUID:99096692; PMID:9880334
A:Accession: T30330
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1087 <STO>
A:Cross-references: EMBL:U95159; NID:94100185; PID:94100186; PID:AA00774.1
C:Genetics:
A:Introns: 137/1

Query Match 15.2%; Score 108; DB 2; Length 1087;
Best Local Similarity 27.1%; Pred. No. 12;
Matches 39; Conservative 28; Mismatches 45; Indels 32; Gaps 6;

QY 19 DQILEYKQVSLKEKAILQATLR-----QATLRREKKLRVENAKLKEIEBELKQELIQAEIO 72
DB 708 DPLIST-KQLRLKQKEDLEKQKQVDSLSQKVSQKQKQKQKQKQKQKQKQKQKQKQ 763
QY 73 NGVQIAPFSGTGPLHANSWSENVIQSTAVTTSSGRTQIKGTGD-----EK 121

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:19:07 ; Search time 16.7176 Seconds
(without alignments)
372.044 Million cell updates/sec

Title: US-09-930-169-1

Perfect score: 709
Sequence: 1 MANNDAVLKRLQKGAERADQ.....EKKGKKKKQKQSIAGSADS 147

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents AA*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	704	99.3	312	3	US-08-360-821B-36 Sequence 36, Appli
2	650	91.7	310	1	US-08-129-456A-37 Sequence 37, Appli
3	509	71.8	310	1	US-08-129-456A-36 Sequence 36, Appli
4	509	71.8	310	2	US-08-705-868-3 Sequence 3, Appli
5	509	71.8	310	3	US-09-123-615-3 Sequence 3, Appli
6	509	71.8	310	3	US-08-360-821B-35 Sequence 35, Appli
7	109	15.4	461	1	US-08-186-222-2 Sequence 2, Appli
8	106.5	15.0	414	5	PCT-US93-03077-3 Sequence 3, Appli
9	106.5	15.0	1093	5	PCT-US93-03077-1 Sequence 1, Appli
10	104	14.7	961	4	US-09-914-259-66 Sequence 66, Appli
11	102	14.4	290	2	US-08-903-801-1 Sequence 1, Appli
12	102	14.4	290	3	US-09-295-055-1 Sequence 1, Appli
13	102	14.4	431	4	US-09-286-981B-3 Sequence 3, Appli
14	101.5	14.3	271	1	US-08-664-596B-11 Sequence 11, Appli
15	101	14.2	956	4	US-09-914-259-17 Sequence 16, Appli
16	101	14.2	957	4	US-09-914-259-16 Sequence 15, Appli
17	100.5	14.2	419	4	US-09-286-981B-15 Sequence 42, Appli
18	100	14.1	588	4	US-08-714-741-42 Sequence 40, Appli
19	100	14.1	864	4	US-08-714-741-40 Sequence 48, Appli
20	98.5	13.9	525	4	US-09-107-532A-5095 Sequence 482, App
21	98	13.8	422	4	US-09-071-035-484 Sequence 482, App
22	98	13.8	449	4	US-09-071-035-482 Sequence 3, Appli
23	97	13.7	234	2	US-08-903-801-3 Sequence 3, Appli
24	97	13.7	234	3	US-09-295-055-3 Sequence 3, Appli
25	96.5	13.6	210	4	US-09-222-938A-67 Sequence 67, Appli
26	95	13.4	262	4	US-09-595-684B-31 Sequence 31, Appli
27	94	13.3	413	4	US-09-286-981B-5 Sequence 5, Appli

28	94	13.3	667	3	US-09-071-709-9 Sequence 9, Appli
29	93.5	13.2	459	3	US-09-071-709-1 Sequence 1, Appli
30	93	13.1	498	4	US-09-107-532A-6991 Sequence 6991, Ap
31	93	13.1	534	4	US-09-103-664A-2 Sequence 2, Appli
32	93	13.1	1388	4	US-08-685-576-1 Sequence 1, Appli
33	92.5	13.0	251	4	US-09-286-981B-4 Sequence 4, Appli
34	92.5	13.0	708	3	US-08-235-836C-76 Sequence 85, Appli
35	92.5	13.0	803	4	US-09-154-750A-85 Sequence 10, Appli
36	92	13.0	414	4	US-09-286-981B-10 Sequence 6, Appli
37	92	13.0	564	3	US-09-308-022-6 Sequence 25, Appli
38	92	13.0	631	3	US-08-847-065-25 Sequence 60, Appli
39	92	13.0	741	4	US-09-854-856-60 Sequence 44, Appli
40	92	13.0	769	4	US-09-854-856-44 Sequence 28, Appli
41	92	13.0	801	4	US-09-854-856-28 Sequence 12, Appli
42	92	13.0	829	4	US-09-854-856-12 Sequence 54, Appli
43	92	13.0	894	4	US-09-854-856-54 Sequence 38, Appli
44	92	13.0	922	4	US-09-854-856-38 Sequence 22, Appli
45	92	13.0	954	4	US-09-854-856-22 Sequence 22, Appli

ALIGNMENTS

RESULT 1
US-08-360-821B-36
Sequence 36, Application US/08360821B
Patent No. 6228837
GENERAL INFORMATION:
APPLICANT: Stern, David M.
APPLICANT: Claus, Matthias
APPLICANT: Kao, Janet
APPLICANT: Kayton, Mark
APPLICANT: Libutti, Steven K
TITLE OF INVENTION: Endothelial Monocyte Activating
TITLE OF INVENTION: Polypeptide II: A Mediator Which Activates Host Response
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESS: Cooper & Dunham, LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.30, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/360, 821B
FILING DATE: 08-OCT-96
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41735-A-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 312 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-360-821B-36
Query Match 99.3%; Score 704; DB 3; Length 312;
Best Local Similarity 99.3%; Pred. No. 1.8e-58;
Matches 146; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Cy 1 MANNDAVLKRLQKGAERADQIEYLKQOVSLKKEKAILQATLREKKLRVENAKKKEIE 60

Db 1 MANNDAVLKRLBQKGAADQIIETLKQVSLKKEKALQATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIQAIEIONGVQIAFPSSGTPPLHANSMSENVIOSTAVTTSSTGTEQIKGCTGD 120
Db 61 ELKQELIQAIEIONGVQIIPPSGTPPLHANSMSENVIOSTAVTTSSTGTEQIKGCTGD 120
QY 121 KKAKEKIEKKEKKEKKQOSIAGSADS 147
Db 121 KKAKEKIEKKEKKEKKQOSIAGSADS 147

RESULT 2

US-08-129-456A-37
; Sequence 37, Application US/08129456A
; Patent No. 5641867
; GENERAL INFORMATION:
; APPLICANT: Stern, David M.
; APPLICANT: Claus, Mathias
; APPLICANT: Kao, Janet
; APPLICANT: Kayton, Mark
; APPLICANT: Libutti, Steven K.
; TITLE OF INVENTION: Endothelial-Monocyte Activating
; TITLE OF INVENTION: Polypeptide II: A Mediator Which
; TITLE OF INVENTION: Activates Host Response
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0 Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/129,456A
; FILING DATE: 29-SEP-1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41735
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 278 0400
; TELEFAX: 212 391 0526
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 310 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-129-456A-37

Query Match 91.7%; Score 650; DB 1; Length 310;
Best Local Similarity 95.9%; Pred. No. 2,1e-53;
Matches 142; Conservative 0; Mismatches 2; Indels 4; Gaps 3;
QY 1 MANNDAVLKRLBQKGAADQIIETLKQVSLKKEKALQATLREKKLRVENAKLKEIE 60
Db 1 MANNDAVLKRLBQKGAADQIIETLKQVSLKKEKALQATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIQAIEIONGVQIAFPSSGTPPLHANSMSENVIOSTAVTTSSTGTEQIKGCTGD 119
Db 61 ELKQELIQAIEIONGVQIIPPSGTPPLHANSMSENVIOSTAVTTSSTGTEQIK--AGD 118
QY 120 EKAKEKIEKKEKKEKKQOSIAGSADS 147
Db 119 EKAKEKIEKKEKKEKKQOSIAGSADS 145

RESULT 3

US-08-129-456A-36
; Sequence 36, Application US/08129456A
; Patent No. 5641867
; GENERAL INFORMATION:
; APPLICANT: Stern, David M.
; APPLICANT: Claus, Mathias
; APPLICANT: Kao, Janet
; APPLICANT: Kayton, Mark
; APPLICANT: Libutti, Steven K.
; TITLE OF INVENTION: Endothelial-Monocyte Activating
; TITLE OF INVENTION: Polypeptide II: A Mediator Which
; TITLE OF INVENTION: Activates Host Response
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0 Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/129,456A
; FILING DATE: 29-SEP-1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41735
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 278 0400
; TELEFAX: 212 391 0526
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 310 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-129-456A-36

Query Match 71.8%; Score 509; DB 1; Length 310;
Best Local Similarity 76.4%; Pred. No. 3,5e-40;
Matches 113; Conservative 8; Mismatches 23; Indels 4; Gaps 3;
QY 1 MANNDAVLKRLBQKGAADQIIETLKQVSLKKEKALQATLREKKLRVENAKLKEIE 60
Db 1 MANNDAVLKRLBQKGAADQIIETLKQVSLKKEKALQATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIQAIEIONGVQIAFPSSGTPPLHANSMSENVIOSTAVTTSSTGTEQIKGCTGD 119
Db 61 ELKQELIQAIEIONGVQVAVRSTPLQCTHASESVQSPVATTSPTKQIK--AGE 118
QY 120 EKAKEKIEKKEKKEKKQOSIAGSADS 147
Db 119 EKAKEKIEKKEKKEKKQOSIAGSADS 145

RESULT 4

US-08-705-868-3
; Sequence 3, Application US/08705868
; Patent No. 5885798
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Coleman, Roger
; APPLICANT: Au-Young, Janice
; APPLICANT: Murty, Lynn B.
; TITLE OF INVENTION: NOVEL MONOCYTE ACTIVATING CYTOKINE

NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/705, 868
FILING DATE: Filed Herewith
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0117 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 498910

Query Match	71.8%	Score 509;	DB 2;	Length 310;
Best Local Similarity	76.4%;	Pred. No. 3.5e-40;		
Matches 113; Conservative	8;	Mismatches 23;	Indels 4;	Gaps 3;

QY	1	MAANDAVYKRLBQKGEADQIIIEYLKQVSLKKEKALLOATLREEKKLRENAKLKEKE	60
Db	1	MAANDAVYKRLBQKGEADQIIIEYLKQVALLKEKALLOATMKEKKLRENAKLKEKE	60
QY	61	ELKQELIOAEINQVQKQIAFPSTGPLHANSWSENATOSTAV--TVWSGTHQEIQKSGTD	119
Db	61	ELKQELIOAEINQVQVRLSTPLQTNCTASVSQVSPVATYASPAHQEIK--AGE	118
QY	120	EKKAKEKIEKKEKKEKQKQSIAGSDS	147
Db	119	EKKAKEKIEKKEKKE--KQKQSPAASTDS	145

RESULT 5
 US-09-123-615-3
 : Sequence 3, Application US/09123615
 : Patent No. 6090377
 : GENERAL INFORMATION:
 : APPLICANT: Bandman, Olga
 : APPLICANT: Coleman, Roger
 : APPLICANT: Au-Young, Janice
 : APPLICANT: Murry, Lynn E.
 : TITLE OF INVENTION: NOVEL MONOCYTE ACTIVATING CYTOKINE
 : NUMBER OF SEQUENCES: 4
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Incyte Pharmaceuticals, Inc.
 : STREET: 3174 Porter Drive
 : CITY: Palo Alto
 : STATE: CA
 : COUNTRY: U.S.
 : ZIP: 94304
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette
 : COMPUTER: IBM Compatible

```

1 OPERATING SYSTEM: DOS
2 SOFTWARE: FastSeq Version 1.5
3 CURRENT APPLICATION DATA:
4 APPLICATION NUMBER: US/09/123,615
5 FILING DATE:
6 PRIOR APPLICATION DATA:
7 APPLICATION NUMBER: 08/705,868
8 FILING DATE:
9 ATTORNEY/AGENT INFORMATION:
10 NAME: Billings, Lucy J.
11 REGISTRATION NUMBER: 36,749
12 REFERENCE/DOCKET NUMBER: PF-0117 US
13 TELECOMMUNICATION INFORMATION:
14 TELEPHONE: 415-855-0555
15 TELEFAX: 415-845-4166
16 INFORMATION FOR SEQ ID NO: 3:
17 SEQUENCE CHARACTERISTICS:
18 LENGTH: 310 amino acids
19 TYPE: amino acid
20 STRANDEDNESS: single
21 TOPOLOGY: linear
22 MOLECULE TYPE: peptide
23 IMMEDIATE SOURCE:
24 LIBRARY: GenBank
25 CLONE: 498910
26
27 US-09-123-615-3

```

Query Match	71.8%;	Score 509;	DB 3;	Length 310;
Best Local Similarity	76.4%;	Pred. No. 3.5e-40;		
Matches 113; Conservative	8;	Mismatches 23;	Indels 4;	Gaps 3

Qy	1	MAANDAVLRLRLOKAGENDQIIIEYKQOVSLLKKEKALQALREKTLRYENNAUKKEIE	60
Db	1	MAINDAVLRRLRLOKAGENDQIIIEYKQOVSLLKKEKALQALREKTLRYENNAUKKEIE	60
Qy	61	ELKOEELIOWEIONGYKOIAFPBSGTPFLHANSVSENVIOSTAV--TTVSSGTEOKIGSTGD	119
Db	61	ELKOEELIOWEIONGYKOIAFPBSGTPFLHANSVSENVIOSTAV--TTVSSGTEOKIGSTGD	118

QY 120 EKKAKEKEKEKEKKQOSIAGSADS 147
Db 119 EKKVKEKEKEKEKKKE-KQOSAASTDS 145

RESULT 6
 US-08-360-821B-35
 Sequence 35, Application US/08360821B
 Patent No. 6228837
 GENERAL INFORMATION:
 APPLICANT: Stern, David M.
 APPLICANT: Clausen, Mathias
 APPLICANT: Kao, Janet
 APPLICANT: Kayton, Mark
 APPLICANT: Libutti, Steven K
 TITLE OF INVENTION: Endothelial Monocyte Activating
 Peptide: Polypeptide II: A Mediator Which Activates Host Responses
 NUMBER OF SEQUENCES: 42
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Cooper & Dunham, LLP
 STREET: 1185 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.30, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/360,821B
 FILING DATE: 08-OCT-96
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41735-A-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-360-821B-35

Query Match 71.8%; Score 509; DB 3; Length 310;
Best Local Similarity 76.4%; Pred. No. 3,5e-40;
Matches 113; Conservative 8; Mismatches 23; Indels 4; Gaps 3;

QY 1 MANNDAVLRLEQKGAADQIIIEYLKQVSLKEKALIQATLREKKLRVENAKLKEE 60
DB 1 MANNDAVLRLEQKGAADQIIIEYLKQVSLKEKALIQATLREKKLRVENAKLKEE 60

QY 61 ELKQELIOAEIONGVQIAFPSPGTPPLHANSWSENVIQSTAV-TTVSSGTKEQIKGTGD 119
DB 61 ELKQELIOAEIONGVQIAFPSPGTPPLHANSWSENVIQSTAV-TTVSSGTKEQIKGTGD 119

QY 120 EKKAKERIEKKGKKEKKEKQOSTGSDS 147
DB 119 EKKVKEKTEKKGKKEKKEKQOSTGSDS 145

RESULT 7
US-08-186-222-2
Sequence 2, Application US/08186222
Patent No. 5559007
GENERAL INFORMATION:
APPLICANT: Surti, Bruno
TITLE OF INVENTION: Bacterial Vectors
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08186,222
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,205
FILING DATE: 19-MAR-1991
APPLICATION NUMBER: GB 9006400.7
FILING DATE: 22-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Villamilzar, JoAnn
REGISTRATION NUMBER: 30,598
REFERENCE/DOCKET NUMBER: 4-17994/A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914)785-7121
TELEFAX: (914)347-5769
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 461 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-186-222-2

Query Match 15.4%; Score 109; DB 1; Length 461;
Best Local Similarity 26.5%; Pred. No. 0.019;
Matches 41; Conservative 33; Mismatches 69; Indels 12; Gaps 5;

QY 1 MANNDAVLRLEQKGAADQIIIEYLKQVSLKEKALIQATLREKKLRVENAKL 55
DB 33 IAKQDATTSSAQSAQAQAVDSLSQKSTTKAQIAKESERKALMQIATL 92

QY 56 KKEIEELKQELIOAEIONGVQIAFPSPGTPPLHANSWSENVIQSTAV-TTVSSGTKEQIK 114
DB 93 NESIKE-RTKLEAQARSQAVNSAATNVDVAVNSKSLTDVQKTAIVTSANQMLE 151

QY 115 GGTDEKAKAKIEKKGKKEKQO--STAGSDS 147
DB 152 ---QOEKEQELSKSETVKKNVNFVSLQSGLDS 183

RESULT 8
PCT-US93-03077-3
Sequence 3, Application PC/TUS9303077
GENERAL INFORMATION:
APPLICANT: Board of Regents, The University of Texas System
APPLICANT: Gaynor, Richard B.
TITLE OF INVENTION: PROTEIN CELLULAR FACTOR USEFUL FOR
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/03077
FILING DATE: 19930331
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/862,025
FILING DATE: April 2, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Kammerer, Patricia A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: UTPD270PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-787-1540
TELEFAX: 713-749-2679
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 414 amino acids
TYPE: AMINO ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
PCT-US93-03077-3

Query Match 15.0%; Score 106.5; DB 5; Length 414;
Best Local Similarity 22.4%; Pred. No. 0.029;
Matches 37; Conservative 33; Mismatches 50; Indels 45; Gaps 5;

QY 3 MNDVAVLRLEQKGAADQIIIEYLKQVSLKEKALIQATLREKKLRVENAKLKK 57

```

Db 1638 NKIRAEQKKEEKEEKKAEELKKEEKEEKKKA-EQLKKEEKEEKKVQQLKKEEKEEKKK 1696
QY 64 QELIQAEIQNGVQIAPPSGTPPLANSVSENVIOSTAVTTVSSGTGEQIKGTGDEKKA 123
Db 1697 AEQKKEE-----ENKIK-----VQQLKKEEKEEKKK 1724
QY 124 KEKIEKKGEKKKKQ 139
Db 1725 AEELKKEEKEEKKVQ 1740

RESULT 10
Q8IHP3 PRELIMINARY; PRT; 2055 AA.
AC Q8IHP3
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE MAEBL, putative.
GN PF11_0486.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=1236864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.D., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AE014843; AAN36066.1; --
SQ SEQUENCE 2055 AA; 243210 MW; 783F48BB7286B1F3 CRC64;

Query Match 15.7%; Score 111.5; DB 5; Length 2055;
Best Local Similarity 28.7%; Pred. No. 27;
Matches 39; Conservative 24; Mismatches 40; Indels 33; Gaps 3;

QY 4 NDAVLKRLQKGAADQIIIEYLKQVSLKKEKAILQATLREKKLRVENAKLKKEIEELK 63
Db 1638 NKIRAEQKKEEKEEKKAEELKKEEKEEKKKA-EQLKKEEKEEKKVQQLKKEEKEEKKK 1696
QY 64 QELIQAEIQNGVQIAPPSGTPPLANSVSENVIOSTAVTTVSSGTGEQIKGTGDEKKA 123
Db 1697 AEQKKEE-----ENKIK-----VQQLKKEEKEEKKK 1724
QY 124 KEKIEKKGEKKKKQ 139
Db 1725 AEELKKEEKEEKKVQ 1740

RESULT 11
Q8IIG7 PRELIMINARY; PRT; 964 AA.
AC Q8IIG7
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF11_0207.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=1236864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.D., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AE014838; AAN35791.1; --
KW Hypothetical protein.
SQ SEQUENCE 964 AA; 116348 MW; 7EAC615834ED3B9 CRC64;

Query Match 15.7%; Score 111; DB 5; Length 964;
Best Local Similarity 25.0%; Pred. No. 13;
Matches 35; Conservative 43; Mismatches 42; Indels 20; Gaps 7;

QY 8 LKRLQKGAADQIIIEYLKQVSLKKEKAILQATLRE-KGLRVENAKLKKEIEELKOE 65
Db 504 IKEIKKEIKKEVEIKKEIKKEIKKEIKVEE--IKEIKKEIKKEIKVEIKKEIKVEIKKEIK 561
QY 66 L--IQAEIQNGVQIAPPSGTPPLANSVSENVIO--STAVTTVSSGTGEQIKGTGDEK 121
Db 562 IKEIKKEIKKEIKVEIKKEIKVEE-----KEIKKEIKKEIKVEIKKEIKVEIKKEIK 611
QY 122 KAKEKI--EKKGEKKKKQ 139
Db 612 EVKEIEKEVEIKKEIKVEE 631

RESULT 12
Q9JH24 PRELIMINARY; PRT; 837 AA.
AC Q9JH24
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE GRIP-associated protein 1 long form.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=2035053; PubMed=10896157;
RA Ye B., Liao D., Zhang X., Zhang P., Dong H., Haganir R.L.;
RT "GRASP-1: A Neuronal RasGEF Associated with the AMPA Receptor/GRIP
RT Complex."
RL Neuron 26:603-617(2000).
DR EMBL; AF274057; AAF82298.1; --
SQ SEQUENCE 837 AA; 96074 MW; A746AE4FD09D3AD2 CRC64;

Query Match 15.6%; Score 110.5; DB 11; Length 837;
Best Local Similarity 22.5%; Pred. No. 12;
Matches 47; Conservative 29; Mismatches 62; Indels 71; Gaps 6;

QY 1 MANDAVLKRLQKGAADQIIIEYLKQVSLKKEKAILQATLRE-EKK 47
Db 31 LRKRGVSLIRQKVAVLDR--EFSKQKALSKSKQAQVEVLLSEKEMLOAKLHSQED 88
QY 48 LRVENAKLKKEIEELKQELIQAEIQNGVQIAPPSGTPPLANSVSENVIOSTAVTTVSS 107
Db 89 FRLQNSTLMAEFSKLSQELQELQAEIQNGVQIAPPSGTPPLANSVSENVIOSTAVTTVSS 148
QY 108 GTRQEQ-----IKGTGDE-----EKK 122

```

DB 149 ALGORYCAVRSVSEGGDPPGVDLPISLSPMLAEVLKWMEREKLLIMQLQ 208
 QY 123 -----AKKIEKKKKKKQOS 140
 DB 209 LESSKQAEISRLQELAKLSEKKKKQOS 237

RESULT 13

ID P90895 PRELIMINARY; PRT; 947 AA.

AC P90895; Q18791;
 DT 01-MAY-1997 (TEMBLrel. 03, Created)
 DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DE F57F5.2 protein.

GN F57F5.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peioderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;

RP SEQUENCE FROM N.A.

RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; Z78012; CAB01418.1; -.

DR EMBL; Z75953; CAB00103.1; JOINED.

DR EMBL; Z75953; CAB00103.1; -.

DR EMBL; Z78012; CAB00103.1; JOINED.

DR HSSP; P26769; 1AB8.

DR WormPep; F57F5.2; CE20887.

DR InterPro; IPR001054; G_cyclase.

DR Pfam; PF00211; guanylate_cyc; 1.

DR SMART; SM00044; CYCC; 1.

DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.

DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.

KW Lyase.

SQ SEQUENCE 947 AA; 106998 MW; 947968427B146D64 CRC64;

Query Match 15.4%; Score 110; DB 5; Length 947;

Best Local Similarity 21.6%; Pred. No. 15;

Matches 37; Conservative 43; Mismatches 59; Indels 32; Gaps 6;

QY 1 MANNDVATKLE---QKGAEDQIIEYLKQVSLKEKAILQATLREKKLRVENAKLKK 57

DB 741 LSSSDLAVSRTSKDSDGTPRTSELKVRIRREALAQEK-EERTTKENQKIEE 799

QY 58 EIEELKQE---LIQAEIQNGVKQIAF-----PQGIPLHANSVSENVIO 98

DB 800 VGEHDVSEATSLDSEVSHDNNISFQMPSPDSIPHEDRTSLSPATPSEIGDAISKKLE 859

QY 99 -----STAVTVSSGTEKQIKGCTDEKKAKIEKKKKKKQOSIAGS 145

DB 860 KEDNSWSWSSLDERTVSAKPTT-----TRLLNQKLEKKKSSWSS 905

RESULT 14

ID Q9XUK4 PRELIMINARY; PRT; 244 AA.

AC Q9XUK4;
 DT 01-NOV-1999 (TEMBLrel. 12, Created)
 DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DE W07G1.1 protein.

GN W07G1.1.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peioderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RP SEQUENCE FROM N.A.

RA Percy C.M.;

RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;

RA none;

RT "Genome sequence of the nematode C.elegans: A platform for

investigating biology.";

RL Science 282:2012-2018 (1998).

DR EMBL; Z82076; CAB04933.1; -.

DR WormPep; W07G1.1; CE24035.

DR InterPro; IPR003533; DCX.

DR Pfam; PF03607; DCX; 1.

DR PROSITE; PS50309; DC; 1.

SQ SEQUENCE 244 AA; 27832 MW; 2F30B876F76FADIC CRC64;

Query Match 15.4%; Score 109.5; DB 5; Length 244;

Best Local Similarity 24.8%; Pred. No. 4;

Matches 37; Conservative 35; Mismatches 54; Indels 23; Gaps 5;

QY 10 RLEQKGAEDQIIE---YKQVSL-LKEKAILQATLREKKLRVENAKLKKIEE 61

DB 75 KLKKKKKKKQTVQPRTEHIEIGDEISIPLNEPVLEKKKKKKLVQKNLDHGSIE 134

QY 62 LKQEL---IQAEIQNGVKQIAFP---SGTPLHANSVSENVIOSTAVTVSSGTEKQIKG 115

DB 135 LKTAETELPSEIIBSEDEHHRRSATPDHHPATKHSVAPSRASR-----RG 185

QY 116 GTGDEKKAKEKIEKKKKKKQOSIAGS 144

DB 186 SRRDEKREDEBERKKKDEKKKSSVSS 214

RESULT 15

ID O96216 PRELIMINARY; PRT; 2385 AA.

AC O96216;
 DT 01-MAY-1999 (TEMBLrel. 10, Created)
 DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DE Hypothetical protein.

GN PF0615C.

OS Plasmodium falciparum (isolate 3D7).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=36329;

RP SEQUENCE FROM N.A.

RL MEDLINE=99021743; PubMed=9804551;

RA Gardner M.J., Tetteelin H., Carucci D.J., Cummings L.M., Aravind L.,

RA Koonin E.V., Shalimov S., Mason T., Yu K., Fujii C., Pederson J.,

RA Shen K., Jing J., Astor C., Lai Z., Schwartz D.C., Perea M.,

RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,

RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;

RT "Chromosome 2 sequence of the human malaria parasite Plasmodium

falciparum.";

RL Science 282:1126-1132 (1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=3D7;

RX MEDLINE=2255705; PubMed=12368864;

RA Gardner M.J., Hall N., Fung B., White O., Berriman M., Hyman R.W.,

RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,

RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyte S.,

RA Chan M.-S., Nene V., Shalimov S.J., Sub B., Peterson J., Angiolini S.,

RA Perea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,

RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,

RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,

RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,

RA Fraser C.M., Barrall B.;

RT "Genome sequence of the human malaria parasite Plasmodium

falciparum.";

RL Nature 419:498-511 (2002).

DR EMBL; AE001406; AAC71912.2; -.

KW Hypothetical protein.

SQ SEQUENCE 2385 AA; 288299 MW; CC7CF642BCTFE0B9 CRC64;

Db 135 HNSNIIKKLRKAKDKNENMVAKLKKVELEELQHLQVADGKEVEKQRENIKKLNS 194
Qy 58 EIEELKQEL-----IOAEIONGVQIAFPSTPLHANSMTSENVIOSTA 101
Db 195 MVEHQEKDGLQVDMDELEEKNSIOAALDSAYTEL-----TLUHKNAKDSIAQDA 249
Qy 102 VTVSSGTEQIKGDTGDEKAKEX---IEKKGEKKKKQOOSIA 142
Db 250 LSR-----EMKAEELSAALEKAQSEARQOQETLA 279

RESULT 9
PCT-US93-03077-1
Sequence 1, Application PC/TUS9303077
GENERAL INFORMATION:
APPLICANT: Board of Regents, The University of Texas System
APPLICANT: Gaynor, Richard B.
APPLICANT: Wu, Foon Kin
TITLE OF INVENTION: PROTEIN CELLULAR FACTOR USEFUL FOR
TITLE OF INVENTION: REGULATING GENE EXPRESSION
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESS: Arnold White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/03077
FILING DATE: 19930331
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/862,025
FILING DATE: April 2, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Kammerer, Patricia A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: UTPD270PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-787-1540
TELEFAX: 713-749-2679
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1093 amino acids
TYPE: AMINO ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
PCT-US93-03077-1

Query Match 15.0%; Score 106.5; DB 5; Length 1093;
Best Local Similarity 22.4%; Pred. No. 0.099;
Matches 37; Conservative 33; Mismatches 50; Indels 45; Gaps 5;

Qy 3 NNDVAVLRLEKQKGAADQIIETLKQVSLKE-----KALIQATIREKKIARVENAKKK 57
Db 571 HNSNIIKKLRKAKDKNENMVAKLKKVELEELQHLQVADGKEVEKQRENIKKLNS 630
Qy 58 EIEELKQEL-----IOAEIONGVQIAFPSTPLHANSMTSENVIOSTA 101
Db 631 MVEHQEKDGLQVDMDELEEKNSIOAALDSAYTEL-----TLUHKNAKDSIAQDA 685
Qy 102 VTVSSGTEQIKGDTGDEKAKEX---IEKKGEKKKKQOOSIA 142
Db 686 LSR-----EMKAEELSAALEKAQSEARQOQETLA 715

Db 135 HNSNIIKKLRKAKDKNENMVAKLKKVELEELQHLQVADGKEVEKQRENIKKLNS 194
Qy 58 EIEELKQEL-----IOAEIONGVQIAFPSTPLHANSMTSENVIOSTA 101
Db 195 MVEHQEKDGLQVDMDELEEKNSIOAALDSAYTEL-----TLUHKNAKDSIAQDA 249
Qy 102 VTVSSGTEQIKGDTGDEKAKEX---IEKKGEKKKKQOOSIA 142
Db 250 LSR-----EMKAEELSAALEKAQSEARQOQETLA 279

RESULT 10
US-09-914-259-66
Sequence 66, Application US/09914259
Patent No. 6495336
GENERAL INFORMATION:
APPLICANT: Makowski, Lee
APPLICANT: Hyman, Paul
APPLICANT: Williams, Mark
TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
FILE REFERENCE: 8471-010-999
CURRENT APPLICATION NUMBER: US/09/914,259
CURRENT FILING DATE: 2000-11-21
NUMBER OF SEQ ID NOS: 180
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 66
LENGTH: 961
TYPE: PRT
ORGANISM: Bos taurus
US-09-914-259-66

Query Match 14.7%; Score 104; DB 4; Length 961;
Best Local Similarity 23.3%; Pred. No. 0.14;
Matches 42; Conservative 35; Mismatches 45; Indels 58; Gaps 7;

Qy 1 MANDVAVLRLEKQKGAADQIIETLKQVSLKE-----EKALIQ--ATLREKK-----L 48
Db 645 LEQHDSTVTHYKNMIRBDQLLEELKQOISTLKQNEQLOTAIVQVSOIQHDKQYNLL 704
Qy 49 RV-----ENAKLKEIEELK--QELIOAEIONGVQIAFPS 82
Db 705 KVGQKDSQHGCPYTDGQNMNGVQPEISIRREIEELKSRRLQSQALAE----- 755
Qy 83 GTPHANSMTSENVIOSTAVTVSSGTEQIKGSTD- EKKAKEXIEKKGEKKKKQOOSI 141
Db 756 -----KDSLIEMLKKSQSPGTNEQSSATATGDSQIAIEIKQELATIKSQNLQSV 805

RESULT 11
US-08-903-801-1
Sequence 1, Application US/08903801
Patent No. 5932712
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: NEW ANEXIN BINDING PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESS: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/903,801
FILING DATE: Herewith
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0354 US
TELECOMMUNICATION INFORMATION:

APPLICANT: Evans, Cheryl
 APPLICANT: Spaulding, Vikki
 APPLICANT: Bowman, Michael
 TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
 TITLE OF INVENTION: ENCODING THEM
 NUMBER OF SEQUENCES: 37
 CORRESPONDENCE ADDRESS:
 ADDRESS: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/664,596B
 FILING DATE:
 CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:
 NAME: Brown, Scott A.
 REGISTRATION NUMBER: 32,724
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 498-8224
 TELEFAX: (617) 876-5851
 INFORMATION FOR SEQ ID NO: 11:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 271 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-664-596B-11

Query Match 14.3%; Score 101.5; DB 1; Length 271;
 Best Local Similarity 27.7%; Pred. No. 0.05;
 Matches 39; Conservative 21; Mismatches 48; Indels 33; Gaps 7;
 QY 1 MANDAVLRLEQKGAEDQIIIEYKQVSLKEKAILQATLREKKLRVENAKLKEIE 60
 Db 131 MENS-----LEKGA-----EINPKTALLPSDSVF-----AERNLKGLE 168
 QY 61 ELKQELIQAEIQNGKQIAFPSTPLHNSWSENVIOSTAVTYVSSGTEQIKGTDE 120
 Db 169 ESPSE--AEDPISGITQTMVAEVAEVEKNETVSB-ILPSTCIVTLVPGIP-----TGDE 219
 QY 121 KKAKEK--IEKKGEKKKKQ 139
 Db 220 KTVDKKNIKKKGNMDEKEK 240

RESULT 15
 US-09-914-259-17
 Sequence 17, Application US/09914259
 Patent No. 6495336
 GENERAL INFORMATION:
 APPLICANT: Makowski, Lee
 APPLICANT: Hyman, Paul
 APPLICANT: Williams, Mark
 TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
 FILE REFERENCE: 8471-010-999
 CURRENT APPLICATION NUMBER: US/09/914,259
 NUMBER OF SEQ ID NOS: 180
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 17
 LENGTH: 956
 TYPE: PRT
 ORGANISM: Mus musculus
 US-09-914-259-17

Query Match 14.2%; Score 101; DB 4; Length 956;
 Best Local Similarity 25.0%; Pred. No. 0.27;
 Matches 33; Conservative 30; Mismatches 55; Indels 14; Gaps 4;
 QY 6 AVLRLEQKGAED---QIIIEYKQVSLKEKAILQATLREKK-----LIVENAKL 55
 Db 422 SLVRQLDDKDEINQSQSLAEKIKQ--MLDQDELLASTRDYKIQBELTRLQIENEA 479
 QY 56 KKEIEELKQELIQAEIQNGKQIAFPSTPLHNSWSENVIOSTAVTYVSSGTEQIKG 115
 Db 480 KDEVEVLQALEELAVNYDQKSGVEDKT--RANEQLTDELAQKTTTLTTQRELSQL 537
 QY 116 GTGDEKKAKEXI 127
 Db 538 LSNHKKRATFI 549

Search completed: January 2, 2004, 18:23:40
 Job time : 16.7176 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:22:28 / Search time 32.8588 Seconds
(without alignments)
897.605 Million cell updates/sec

Title: US-09-930-169-1

Sequence: 1 MANNDAVILKRLKQKAEADQ.....EKKEKKKKQKQSIAGSADS 147

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 733937

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubppa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep.*
- 6: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/1/pubppa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	709	100.0	147	11	US-09-930-169-1
2	704	99.3	312	10	US-09-851-026-36
3	516	72.8	108	11	US-09-930-169-2
4	509	71.8	310	10	US-09-851-026-35
5	275	38.8	166	11	US-09-930-169-3
6	147	20.7	58	10	US-09-813-718-22
7	116.5	16.4	1079	12	US-10-369-493-2116
8	107	15.1	1965	12	US-10-369-493-3279
9	106.5	15.0	1069	15	US-10-146-473-77
10	105.5	14.9	1884	10	US-09-785-770A-17
11	105.5	14.9	1907	10	US-09-785-770A-16
12	105	14.8	670	11	US-09-298-523B-63
13	104.5	14.7	1231	12	US-10-369-493-3503
14	104	14.7	711	11	US-09-298-523B-3
15	104	14.7	961	12	US-10-080-608A-66

16	104	14.7	961	12	US-10-370-685-155	Sequence 155, App
17	103.5	14.6	496	11	US-09-298-523B-4	Sequence 4, Appl
18	103	14.5	286	10	US-09-881-752A-352	Sequence 352, App
19	103	14.5	1130	12	US-10-369-493-6751	Sequence 6751, App
20	103	14.5	1169	12	US-10-369-493-1095	Sequence 1095, App
21	102	14.4	290	10	US-09-808-885-1	Sequence 1, Appl
22	102	14.4	431	12	US-10-254-995-3	Sequence 3, Appl
23	101.5	14.3	319	14	US-10-114-893-90	Sequence 17, Appl
24	101.5	14.3	890	12	US-09-848-602-19	Sequence 106, App
25	101	14.2	956	12	US-10-080-608A-17	Sequence 16, App
26	101	14.2	957	12	US-10-370-685-106	Sequence 105, App
27	101	14.2	957	12	US-10-370-685-105	Sequence 15, App
28	100.5	14.2	419	12	US-10-254-995-15	Sequence 33, Appl
29	100.5	14.2	433	11	US-09-056-019-33	Sequence 61, Appl
30	100.5	14.2	690	11	US-09-228-523B-61	Sequence 62, Appl
31	100.5	14.2	701	11	US-09-228-523B-62	Sequence 14, Appl
32	100.5	14.2	589	11	US-09-228-523B-14	Sequence 2, Appl
33	100	14.1	691	11	US-09-228-523B-1	Sequence 1, Appl
34	100	14.1	929	11	US-09-298-523B-2	Sequence 14, Appl
35	100	14.1	707	11	US-09-298-523B-60	Sequence 2, Appl
36	100	14.1	886	12	US-10-369-493-1016	Sequence 60, Appl
37	99.5	14.0	886	12	US-09-727-384-6	Sequence 1016, App
38	99	14.0	2139	9	US-10-023-219-4	Sequence 4, Appl
39	98.5	13.9	1179	12	US-10-369-493-1330	Sequence 1330, App
40	98.5	13.9	1179	12	US-10-369-493-20394	Sequence 20394, App
41	98.5	13.9	1827	12	US-10-369-493-5368	Sequence 5368, App
42	98.5	13.9	1827	12	US-09-925-302-711	Sequence 711, App
43	98	13.8	1203	15	US-10-097-340-43	Sequence 43, Appl
44	98	13.8	1164	12	US-10-369-493-9770	Sequence 9770, App
45	97.5	13.8	1164	12	US-10-369-493-9770	Sequence 9770, App

ALIGNMENTS

RESULT 1
US-09-930-169-1
Sequence 1, Application US/09930169
Publication No. US20030004309A1
GENERAL INFORMATION:
APPLICANT: KIM, SUNGHOON
APPLICANT: KO, YOUNG-GU
TITLE OF INVENTION: IMMUNOLOGICAL ENHANCEMENT AGENT COMPRISING N-TERMINAL PEPTIDE OF
FILE REFERENCE: 058333/0106
CURRENT APPLICATION NUMBER: US/09/930,169
PRIOR FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: KR 2001-31310
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 147
TYPE: PRT
ORGANISM: Unknown Organism
FEATURES:
OTHER INFORMATION: Description of Unknown Organism: Mammalian
US-09-930-169-1

this case

Query Match	100.0%	Score 709	DB 11	Length 147
Best Local Similarity	100.0%	Pred. No. 7.7e-54		
Matches 147	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MANNDAVILKRLKQKAEADQIIIEYIKQVSLIKKATIIQTLREKKLRVENATLKEIE	60	
DB	1	MANNDAVILKRLKQKAEADQIIIEYIKQVSLIKKATIIQTLREKKLRVENATLKEIE	60	
QY	61	EKKELIAPETONGKQIAPSGPTPLANSVSNVISTAVTVSSGTEQIKGIGDE	120	
DB	61	EKKELIAPETONGKQIAPSGPTPLANSVSNVISTAVTVSSGTEQIKGIGDE	120	
QY	121	KKAKKIKKKEKKKKQKQSIAGSADS	147	

Db 121 KKAKETKEKKEKKQSIAGSADS 147

RESULT 2

US-09-851-026-36
Sequence 36, Application US/09851026
Patent No. US20020160957A1

GENERAL INFORMATION:

APPLICANT: Stern, David M.
Clauses, Matthias

Kao, Janet

Libutti, Steven K

TITLE OF INVENTION: Endothelial Monocyte Activating
Polypeptide II: A Mediator Which Activates Host Response

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham, LLP

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.30, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/851,026

FILING DATE: 07-May-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/360,821

FILING DATE: 08-OCT-96

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 41735-A-PCT-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-278-0400

TELEFAX: 212-391-0525

INFORMATION FOR SEQ ID NO: 36:

SEQUENCE CHARACTERISTICS:

LENGTH: 312 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 36:

US-09-851-026-36

Query Match 99.3%; Score 704; DB 10; Length 312;

Best Local Similarity 99.3%; Pred. No. 5.3e-53;

Matches 146; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 MANDAVLKLRLKLEKGAADQIIIEYLKQVSLKEKAILOATLREKKLRVENAKLKEIE 60

1 MANDAVLKLRLKLEKGAADQIIIEYLKQVSLKEKAILOATLREKKLRVENAKLKEIE 60

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSGTRKQIKSGTGE 120

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSGTRKQIKSGTGE 120

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSGTRKQIKSGTGE 120

121 KKAKETKEKKEKKQSIAGSADS 147

121 KKAKETKEKKEKKQSIAGSADS 147

121 KKAKETKEKKEKKQSIAGSADS 147

121 KKAKETKEKKEKKQSIAGSADS 147

121 KKAKETKEKKEKKQSIAGSADS 147

121 KKAKETKEKKEKKQSIAGSADS 147

121 KKAKETKEKKEKKQSIAGSADS 147

121 KKAKETKEKKEKKQSIAGSADS 147

121 KKAKETKEKKEKKQSIAGSADS 147

121 KKAKETKEKKEKKQSIAGSADS 147

121 KKAKETKEKKEKKQSIAGSADS 147

Publication No. US20030004309A1

GENERAL INFORMATION:

APPLICANT: KIM, SUNGHOON

TITLE OF INVENTION: IMMUNOLOGICAL ENHANCEMENT AGENT COMPRISING N-TERMINAL PEPTIDE OF

FILE REFERENCE: 058333/0106

CURRENT FILING DATE: 2001-08-16

PRIOR FILING DATE: 2001-06-05

NUMBER OF SEQ ID NOS: 11

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 2

LENGTH: 108

TYPE: PRT

ORGANISM: Unknown Organism

OTHER INFORMATION: Description of Unknown Organism: Mammalian

OTHER INFORMATION: protein sequence

US-09-930-169-2

Query Match 72.8%; Score 516; DB 11; Length 108;

Best Local Similarity 100.0%; Pred. No. 2.4e-37;

Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MANDAVLKLRLKLEKGAADQIIIEYLKQVSLKEKAILOATLREKKLRVENAKLKEIE 60

1 MANDAVLKLRLKLEKGAADQIIIEYLKQVSLKEKAILOATLREKKLRVENAKLKEIE 60

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPFSGTPELHANSVSENVIOSTAVTVSSG 108

TELERAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 35:
US-09-851-026-35

Query Match 71.8%; Score 509; DB 10; Length 310;
Best Local Similarity 76.4%; Pred. No. 3,66-36;
Matches 113; Conservative 8; Mismatches 23; Indels 4; Gaps 3;

Qy 1 NANNDAVAKRLKLEQKGAADQIIEYKQOVSLKKEKAIQATLREKTLRVENAKLKEIE 60
Db 1 MATNDVAVKRLKLEQKGAADQIIEYKQOVSLKKEKAIQATLREKTLRVENAKLKEIE 60
Qy 61 ELKQELIAEIQNGVQKQIAPPSGTEPLHANSVSENVIOSTAV-TTVSSGTEKQIKGTGD 119
Db 61 ELKQELIAEIQNGVQKQIAPPSGTEPLHANSVSENVIOSTAV-TTVSSGTEKQIKGTGD 118
Qy 120 EKKAKETKEKKEKKEKQKQSIAGSADS 147
Db 119 EKKAKETKEKKEKKEKQKQSIAGSADS 145

RESULT 5
US-09-930-169-3
Sequence 3, Application US/09930169
Publication No. US200300043092A1
GENERAL INFORMATION:
APPLICANT: KIM, SUNGHOO
APPLICANT: KO, YOUNG-GYU
TITLE OF INVENTION: IMMUNOLOGICAL ENHANCEMENT AGENT COMPRISING N-TERMINAL PEPTIDE OF
TITLE OF INVENTION: P43 AS AN EFFECTIVE COMPONENT
FILE REFERENCE: 058333/0106
CURRENT APPLICATION NUMBER: US/09/930,169
CURRENT FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: KR 2001-31310
PRIOR FILING DATE: 2001-06-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 3
LENGTH: 166
TYPE: PRT
ORGANISM: Unknown Organism
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: Mammalian
US-09-930-169-3

Query Match 38.8%; Score 275; DB 11; Length 166;
Best Local Similarity 100.0%; Pred. No. 2,76-16;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 91 MVSENVIOSTAVTVSSGTEKQIKGTGDEKKAKETKEKKEKQKQSIAGSADS 147
Db 1 MVSENVIOSTAVTVSSGTEKQIKGTGDEKKAKETKEKKEKQKQSIAGSADS 57

RESULT 6
US-09-813-718-22
Sequence 22, Application US/09813718
Publication No. US2002018266A1
GENERAL INFORMATION:
APPLICANT: Schimmel, Paul
APPLICANT: Wakasugi, Keisuke
TITLE OF INVENTION: Human Aminoacyl-tRNA Synthetase Polypeptides Useful For
TITLE OF INVENTION: The Regulation of Angiogenesis
FILE REFERENCE: 00-221
CURRENT APPLICATION NUMBER: US/09/813,718

CURRENT FILING DATE: 2001-03-21
NUMBER OF SEQ ID NOS: 58
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 22
LENGTH: 58
TYPE: PRT
ORGANISM: Homo sapiens
US-09-813-718-22

Query Match 20.7%; Score 147; DB 10; Length 58;
Best Local Similarity 100.0%; Pred. No. 8,26-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 118 GDEKAKETKEKKEKKEKQKQSIAGSADS 147
Db 1 GDEKAKETKEKKEKKEKQKQSIAGSADS 30

RESULT 7
US-10-369-493-2116
Sequence 2116, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 2116
LENGTH: 1079
TYPE: PRT
ORGANISM: Schizosaccharomyces pombe
US-10-369-493-2116

Query Match 16.4%; Score 116.5; DB 12; Length 1079;
Best Local Similarity 21.6%; Pred. No. 0,13;
Matches 38; Conservative 41; Mismatches 58; Indels 39; Gaps 5;

Qy 3 NNDVAVKRLKLEQKGAADQIIEYKQOVSLKKEKAIL-----QATLREKTLRVENA 53
Db 246 NNTALQKMLEKRLKLEQKGAADQIIEYKQOVSLKKEKAIL-----QATLREKTLRVENA 305

Qy 54 KTK-----KEIELEKQELIAEIQNGVQKQIAPPSG---TPLHN----- 89
Db 306 EKKAKETKEKKEKKEKQKQSIAGSADS 147

Qy 90 MVSENVIOSTAVTVSSGTEKQIKGTGDEKKAKETKEKKEKQKQSIAGSADS 145
Db 366 SISSGILSESPAISIVDERQK-----DSKDSKVEKETEVEKRENEAEAEA 415

RESULT 8
US-10-369-493-3279
Sequence 3279, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493

;; CURRENT FILING DATE: 2003-02-28
;; PRIOR APPLICATION NUMBER: US 60/360,039
;; PRIOR FILING DATE: 2002-02-21
;; NUMBER OF SEQ ID NOS: 47374
;; SEQ ID NO 3279
;; LENGTH: 1965
;; TYPE: PRT
;; ORGANISM: Neurospora crassa
US-10-369-493-3279

Query Match 15.1%; Score 107; DB 12; Length 1965;
Best Local Similarity 25.0%; Pred. No. 1.8;
Matches 44; Conservative 35; Mismatches 53; Indels 44; Gaps 9;

QY 2 ANNDVLRLEKQKGAEDQIIEYLKQOVSLLKEKALIQATLREKKLRVENAKLKKKEIE-60
DB 1502 AEPEAKADLEER--ADALVKORSEKMTALNKLAESEAKEMKOTODEROKLOAEYD-1558
QY 61 ELKQEL--IQAEIQNGVKQIAPF-----SGTP-----LHANSWV 92
DB 1559 KIQQELALILKAEQSGVSGVGPATPAKPPTAQTAPGAGTGLPDLNLTDQTRRELI 1618
QY 93 SENVIQSTAVTVSSGTKEQIKGGTDEKAKKIE--KKGEKKK--KQOSIA 142
DB 1619 SSNOV---IMGIKSNVKKNI---AAESKKVRELEVVAEYKOKIASAKEQAAA 1668

RESULT 9

US-10-146-473-77
;; Sequence 77, Application US/10146473
;; Publication No. US20030108888A1
;; GENERAL INFORMATION:
;; APPLICANT: Scanlan, Matthew
;; APPLICANT: Scout, Ivan
;; APPLICANT: Stockert, Elisabeth
;; APPLICANT: Gure, Ali
;; APPLICANT: Chen, Yao-Tseeng
;; APPLICANT: Old, Lloyd
;; TITLE OF INVENTION: Breast Cancer Antigens
;; FILE REFERENCE: L00461/70130(JRV)
;; CURRENT APPLICATION NUMBER: US/10/146,473
;; CURRENT FILING DATE: 2002-05-15
;; PRIOR APPLICATION NUMBER: US 60/291,150
;; PRIOR FILING DATE: 2001-05-15
;; NUMBER OF SEQ ID NOS: 82
;; SOFTWARE: Patentin version 3.0
;; SEQ ID NO 77
;; LENGTH: 1069
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: UNSURE
;; LOCATION: (268)..
;; OTHER INFORMATION: X = any amino acid
US-10-146-473-77

Query Match 15.0%; Score 106.5; DB 15; Length 1069;
Best Local Similarity 22.4%; Pred. No. 0.96;
Matches 37; Conservative 33; Mismatches 50; Indels 45; Gaps 5;

QY 3 NNDVLRLEKQKGAEDQIIEYLKQOVSLLKE-----KALIQATLREKKLRVENAKLKK 57
DB 564 HNGNIITIKLPAKQKKNENMAKANKYKELFEELQHILKQVLDKEVEYKQHRENIKKLNS 623
QY 58 EIEELKQEL-----IQAEIQNGVKQIAPFSGTPLHANSWVSENVIGSTA 101
DB 624 MVRERQEDLRLQVMDLEEKRSIQALDSAYKEL-----TDLHKAANAKSEAOEAA 678
QY 102 VTVSSGTKEQIKGGTDEKAKKIE--KKGEKKK--KQOSIA 142
DB 679 LSR-----EMKAKEELSALEKQOEERKQOQETLA 708

RESULT 10

US-09-785-770A-17
;; Sequence 17, Application US/09785770A
;; Patent No. US20020103360A1
;; GENERAL INFORMATION:
;; APPLICANT: Pan, Yang
;; APPLICANT: Barnes, Thomas M.
;; TITLE OF INVENTION: A NOVEL PROTEIN RELATED TO MELANOMA-INHIBITING PROTEIN
;; FILE REFERENCE: 07334-328001
;; CURRENT APPLICATION NUMBER: US/09/785,770A
;; CURRENT FILING DATE: 2001-02-16
;; PRIOR APPLICATION NUMBER: US 09/387,462
;; PRIOR FILING DATE: 1999-09-01
;; PRIOR APPLICATION NUMBER: US 09/145,056
;; PRIOR FILING DATE: 1998-09-01
;; NUMBER OF SEQ ID NOS: 24
;; SOFTWARE: Patentin Ver. 4.0
;; SEQ ID NO 17
;; LENGTH: 1884
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-785-770A-17

Query Match 14.9%; Score 105.5; DB 10; Length 1884;
Best Local Similarity 26.1%; Pred. No. 2.4;
Matches 41; Conservative 30; Mismatches 55; Indels 31; Gaps 6;

QY 16 APADQIIEYLKQOVSLLKEK-----ALIQATLREKK-----KRVENAKLKEIEELK 63
DB 1282 SENKSIIEKLKVIVISNNAFSEFVOIALNEAKLSEKVSCHGRVQEEENARLKKKEQLO 1341
QY 64 QEL-----IQAEIQNGVK-----QIAPFSGTPLHANSWVSENVIGSTAVTVSSGT 109
DB 1342 QEIEDMSKHAELSEQIISFEKQKLEVAL--THKDNIALTLCTIQLNLCESES 1398
QY 110 KEQIKGTGDEKAKKIEKKGEKKKQOSIAGSAD 146
DB 1399 EGQNGKNDSDLANGEV--GGDRNEKKNQKQKMMMD 1433

RESULT 11

US-09-785-770A-16
;; Sequence 16, Application US/09785770A
;; Patent No. US20020103360A1
;; GENERAL INFORMATION:
;; APPLICANT: Pan, Yang
;; APPLICANT: Barnes, Thomas M.
;; TITLE OF INVENTION: A NOVEL PROTEIN RELATED TO MELANOMA-INHIBITING PROTEIN
;; FILE REFERENCE: 07334-328001
;; CURRENT APPLICATION NUMBER: US/09/785,770A
;; CURRENT FILING DATE: 2001-02-16
;; PRIOR APPLICATION NUMBER: US 09/387,462
;; PRIOR FILING DATE: 1999-09-01
;; PRIOR APPLICATION NUMBER: US 09/145,056
;; PRIOR FILING DATE: 1998-09-01
;; NUMBER OF SEQ ID NOS: 24
;; SOFTWARE: Patentin Ver. 4.0
;; SEQ ID NO 16
;; LENGTH: 1907
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-785-770A-16

Query Match 14.9%; Score 105.5; DB 10; Length 1907;
Best Local Similarity 26.1%; Pred. No. 2.4;
Matches 41; Conservative 30; Mismatches 55; Indels 31; Gaps 6;

QY 16 APADQIIEYLKQOVSLLKEK-----ALIQATLREKK-----KRVENAKLKEIEELK 63
DB 1305 SENKSIIEKLKVIVISNNAFSEFVOIALNEAKLSEKVSCHGRVQEEENARLKKKEQLO 1364

QY 64 QEL-----IQAEIONGVK-----QIAFPGSTPLHANSVSENVIOSTAVTTSSGT 109
Db 1365 QEIEWMSKLAHABLSFOIKSPFKSQDLEVAL---THKDNINATNCITQNLIECESSE 1421
QY 110 KEQIKGTGDEKKAEXIEKGEKKKQOSIAGSAD 146
Db 1422 EGQNKGGNDSDLANGEV--GGDRNEKMKNOIKQMD 1456

RESULT 12

US-09-298-523B-63
; Sequence 63, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 670
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-63

Query Match 14.8%; Score 105; DB 11; Length 670;
Best Local Similarity 27.1%; Pred. No. 0.72;
Matches 38; Conservative 27; Mismatches 51; Indels 24; Gaps 5;

QY 12 EOKGAADQIIEYKQOVSLLKEKALIQATREKKLVE---NAKL-KKEIEELKQEL 67
Db 356 EKVVAEAKKVBEAKKAKQDKEEDRRNYPTNTYKTLLEIASDVKKVKALELVKEEAN 415
QY 68 QAEIQNGVKQIAFPGSTPLHANSVSENVIOSTAVTTSSGTKEQIKGTGDEKKAEXI 127
Db 416 ESRNEEKIKQ-----AKEKVESKKAELATRL-----EKIK--TDRKKAEEBA 455
QY 128 EKKGEKKKKQOSIAGSADS 147
Db 456 KRKAEESEKKAELAKQKVD 475

RESULT 13

US-10-369-493-3503
; Sequence 3503, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3503
; LENGTH: 1231
; TYPE: PRT
; ORGANISM: Neurospora crassa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1231)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-3503

Query Match 14.7%; Score 104.5; DB 12; Length 1231;
Best Local Similarity 23.6%; Pred. No. 1.7;
Matches 35; Conservative 28; Mismatches 56; Indels 29; Gaps 3;

QY 23 EYLQOVSLLKEKALIQATREKKLRVNAKLKKEIEELKQEL--IQAEI--EIQNGVK 76
Db 252 DYVQOEKTLQOSADLEKQKQRELESATRLKNEISHLEEDIQRYVAQORDKEIKGGK 311
QY 77 QIAFPGSTPLHANSV-----SENVIOSTAVTTSSGTKEQI 113
Db 312 AQALBEAVKKAELVRLATVVDLKKSSMKEQERRRAGEQNVADLEALAEKTYTKYKI 371
QY 114 KGTGDEKKAEXIEKGEKKKQOSI 141
Db 372 KAKYDAKETLEKQREQEDTKRELLQTL 399

RESULT 14

US-09-298-523B-3
; Sequence 3, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 711
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-3

Query Match 14.7%; Score 104; DB 11; Length 711;
Best Local Similarity 27.0%; Pred. No. 0.95;
Matches 38; Conservative 27; Mismatches 50; Indels 26; Gaps 5;

QY 12 EOKGAADQIIEYKQOVSLLKEKALIQATREKKLVE---NAKLK-KKEIEELKQEL 66
Db 397 EKVVAEAKKVBEAKKAKQDKEEDRRNYPTNTYKTLLEIASDVKKVKALELVKEEBA 456
QY 67 IQAEIQNGVKQIAFPGSTPLHANSVSENVIOSTAVTTSSGTKEQIKGTGDEKKAEXI 126
Db 457 NESRNEEKIKQ-----AKEKVESKKAELATRL-----EKIK--TDRKKAEEBA 495
QY 127 IEKKGEKKKKQOSIAGSADS 147
Db 496 AKRKAEESEKKAELAKQKVD 516

RESULT 15

US-10-080-608A-66
; Sequence 66, Application US/10080608A
; Publication No. US20030198956A1
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STRAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/10/080,608A
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 961
; TYPE: PRT
; ORGANISM: Bos taurus
US-10-080-608A-66

Query Match 14.7%; Score 104; DB 12; Length 961;
 Best Local Similarity 23.3%; Pred. No. 1.4;
 Matches 42; Conservative 35; Mismatches 45; Indels 58; Gaps 7;

```

QY      1 MANNDAVLKRLKLEQKGAADQIIIEYLKQVSLK-----EKALIQ--ATTREKK-----L 48
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      645 LEQHDSIVTHYKRMIREQDILQLEELKQOISTLKCONEOLOTAVTQOVSOIQHKOYNLL 704
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      49 RV-----ENAKLKKEIEELK--OEIIQAEIIONGVKQIAPPS 82
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      705 KVQLGKDSQHQPYYTDGAQNGVQPEHISRLREIEHLEKSNREILLOSQIAE----- 755
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      83 GTPHLANSWVSENVIOSTAVTVSSGTKEQIKGTGD-EKKAKKIEKKGKKKKQOQSI 141
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      756 -----KDSLIENLKSSQLSPGTNEQSSATAGDSQIIEIKQELATLTKSQLSQSV 805
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    
```

Search completed: January 2, 2004, 18:31:21
 Job time : 32.8588 secs


```

FT Region 75..77
FT /note= "beta1 strand"
FT Region 79..85
FT /note= "beta7 strand"
FT Region 90..92
FT /note= "beta8 strand"
FT Region 103..106
FT /note= "beta9 strand"
FT Region 119..123
FT /note= "alpha 2 helix"
FT Region 124..130
FT /note= "alpha 3 helix"
FT Region 132..134
FT /note= "beta10 strand"
FT Region 140..142
FT /note= "beta11 strand"
XX
XX WO200195927-A1.
XX
XX 20-DEC-2001.
XX
XX 14-JUN-2000; 2000WO-KR00630.
XX
XX 14-JUN-2000; 2000WO-KR00630.
XX
XX (IMAG-) IMAGENE CO LTD.
XX
XX Kim S, Ko Y, Kim YS, Jo YJ;
XX
XX WPI; 2002-098017/13.
XX
XX N-PSDB; ABA94640.
XX
XX Novel anti-tumor and anti-angiogenic agent of p43 comprises N-terminal
XX domain and C-terminal domain containing eleven beta-strands forming a
XX structural core and three flanking alpha-helices -
XX
XX Claim 1; Fig 1; 35pp; English.
XX
XX The invention provides an anti-tumor and anti-angiogenic agent of p43
XX consisting of two domains, the N-terminal domain (146 amino acids) and
XX C-terminal domain (166 amino acids) containing 11 beta-strands forming
XX a structural core and 3 flanking alpha-helices. p43 is useful as an anti-
XX tumor and anti-angiogenic agent. p43 and its C-terminal cytokine domain
XX (EMAP II) induce regression of fibrosarcoma in immunocompromised mouse
XX while its N-terminal domain does not. p43 is a potent cytokine as
XX determined by the induction of tumor necrosis factor-alpha (TNF-alpha),
XX interleukin-6 (IL-6), IL-8 and matrix metalloproteinase-9 or by its
XX activity of chemotaxis. The present sequence represents the human p43
XX polypeptide.
XX
XX Sequence 312 AA;
XX
XX Query Match 100.0%; Score 516; DB 23; Length 312;
XX Best Local Similarity 100.0%; Pred. No. 1.7e-45;
XX Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MANNDAVLRKLEKQKGAADQIIIEYLRKQVSLKEKAILQATLRBEKKLRVENAKLKEIE 60
XX DB 1 MANNDAVLRKLEKQKGAADQIIIEYLRKQVSLKEKAILQATLRBEKKLRVENAKLKEIE 60
XX
XX QY 61 ELKQELIQAEIQNGVQKQIAFPSPGTPHANSVMSENVIGSTAVTTVSSG 108
XX DB 61 ELKQELIQAEIQNGVQKQIAFPSPGTPHANSVMSENVIGSTAVTTVSSG 108
XX
XX RESULT 2
XX AAB63244
XX AAB63244 standard; Protein; 328 AA.
XX
XX AAB63244;
XX
XX 26-MAR-2001 (first entry)
XX

```

```

DE Human breast cancer associated antigen protein sequence SEQ ID NO:606.
XX
XX Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
XX cancer associated antigen; cytostatic; cancer vaccine.
XX
XX Homo sapiens.
XX
XX WO200073801-A2.
XX
XX 07-DEC-2000.
XX
XX 26-MAY-2000; 2000WO-US14749.
XX
XX 28-MAY-1999; 99US-0136526.
XX
XX 10-SEP-1999; 99US-0153454.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX
XX Obata Y;
XX
XX WPI; 2001-025274/03.
XX
XX Nucleic acids encoding breast, gastric and prostate cancer associated
XX antigen precursors, useful for diagnosing and treating a condition
XX characterized by expression of an abnormal amount of a protein, e.g.
XX cancer -
XX
XX Example 1; Page 478; 799pp; English.
XX
XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
XX represent nucleotide sequences encoding human breast, gastric and
XX prostate cancer associated antigen precursors (CAAP) respectively.
XX AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
XX represent human breast, gastric and prostate CAAP protein sequence
XX respectively. CAAPs have cytostatic activity and can be used in the
XX production of cancer vaccines. The human CAAP proteins, peptides, nucleic
XX acids or anti-CAAP antibodies are useful for diagnosing and treating a
XX condition characterised by expression of an abnormal amount of a protein,
XX e.g. cancer.
XX
XX Sequence 328 AA;
XX
XX Query Match 100.0%; Score 516; DB 22; Length 328;
XX Best Local Similarity 100.0%; Pred. No. 1.9e-45;
XX Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MANNDAVLRKLEKQKGAADQIIIEYLRKQVSLKEKAILQATLRBEKKLRVENAKLKEIE 60
XX DB 17 MANNDAVLRKLEKQKGAADQIIIEYLRKQVSLKEKAILQATLRBEKKLRVENAKLKEIE 76
XX
XX QY 61 ELKQELIQAEIQNGVQKQIAFPSPGTPHANSVMSENVIGSTAVTTVSSG 108
XX DB 77 ELKQELIQAEIQNGVQKQIAFPSPGTPHANSVMSENVIGSTAVTTVSSG 124
XX
XX RESULT 3
XX AAR72577
XX AAR72577 standard; Protein; 310 AA.
XX
XX AAR72577;
XX
XX 25-MAR-2003 (updated)
XX
XX 29-SEP-1995 (first entry)
XX
XX Mouse EMAP11.
XX
XX EMAP11; endothelial monocyte activating polypeptide II; chemotaxis;
XX inflammation; tissue factor; tumor; cancer; therapy; metha;
XX sarcoma.
XX
XX Mus musculus.
XX
XX WO9509180-A1.
XX

```

XX 06-APR-1995.
 PD 29-SEP-1994; 94WO-US11085.
 XX 29-SEP-1993; 93US-0129456.
 PR 29-SEP-1993; 93US-0129456.
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 PA Klaus M, Kao J, Kayton M, Libutti SK, Stern DM;
 PI WPI, 1995-147389/19.
 XX N-PSDB; AAQ86718.
 DR New endothelial monocyte activating polypeptide II - induces
 PT chemotaxis, inflammation and tissue factor, useful for treating
 PT tumours, also related antibodies, DNA and active fragments
 XX Disclosure; Fig.4; 180pp; English.
 PS A mouse metha sarcoma CDNA library was screened with a probe
 CC based on the N-terminal sequence of mouse EwAPlI. Overlapping
 CC clones were combined to obtain a contiguous full-length sequence
 CC (given in AA086718) encoding a 33 kDa protein (AA072577).
 CC Recombinant EwAPlI was expressed in E. coli.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC
 XX Sequence 310 AA;

Query Match 76.6%; Score 395; DB 16; Length 310;
 Best Local Similarity 77.6%; Pred. No. 7.1e-33;
 Matches 83; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

QY 1 MANDVAVKRLKQGAADQIIIEYVKQVSLKEXALIQATLRREKKLRVENAKLKEIE 60
 DB 1 MATNDVAVKRLKQGAADQIIIEYVKQVSLKEXALIQATLRREKKLRVENAKLKEIE 60
 QY 61 ELKQELIAEIQNGVQKQIAFPSTPLHANSWSENVISTVTTVSS 107
 DB 61 ELKQELIAEIQNGVQKQIAFPSTPLHANSWSENVISTVTTVSS 107

RESULT 4
 ABB59210 ID ABB59210 standard; Protein; 294 AA.
 AC ABB59210;
 XX 26-MAR-2002 (first entry)
 XX Drosophila melanogaster polypeptide SEQ ID NO 4422.
 DE Drosophila melanogaster polypeptide SEQ ID NO 4422.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 XX Drosophila melanogaster.
 OS WO200171042-A2.
 PN 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI, 2001-6556860/75.
 DR N-PSDB; ABL0313.
 XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 XX Disclosure; SEQ ID NO 4422; 21pp + Sequence listing; English.
 PS The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB057737-AB072072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 294 AA;

Query Match 20.5%; Score 106; DB 22; Length 294;
 Best Local Similarity 31.6%; Pred. No. 0.0089;
 Matches 24; Conservative 21; Mismatches 27; Indels 4; Gaps 1;

QY 6 AVLRLEBQGAADQIIIEYVKQVSLKEXALIQATLRREKKLRVENAKLKEIEELKOE 65
 DB 2 ADLQIANSNNERRARALINSIEAEISGIGQ----QLVERQKQELIKENALAKEVBAALAQ 57
 QY 66 LIQAEIQNGVQKQIAFP 81
 DB 58 LVQLELRNGKQIPIVP 73

RESULT 5
 AAB21233 ID AAB21233 standard; Protein; 672 AA.
 AC AAB21233;
 XX 09-MAR-2001 (first entry)
 XX Corn MFP1.
 DE Corn; MAR binding filament-like protein 1; MFP1;
 KM matrix attachment region; MAR; NEMFP1-2; anchor protein.
 XX Zea mays.
 OS WO200061615-A2.
 PN 19-OCT-2000.
 XX 12-APR-2000; 2000WO-US09723.
 PR 12-APR-1999; 99US-0128900.
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 XX Harder PA, Meier I;
 PI WPI; 2000-679464/66.
 DR N-PSDB; AAA95815.
 XX Nucleic acid fragments from tobacco, corn, soybean and rice, encoding
 PT proteins that are homologous to the MAR binding filament-like protein 1
 PT (MFP1), useful for development of novel phenotypes -
 XX
 XX Claim 4; Page 59-61; 62pp; English.
 PS The present sequence is a corn matrix attachment region (MAR)
 CC binding filament-like protein 1 (MFP1). A corn EST sequence was
 CC identified as an MFP1 homologue from clone p0118.chab48r. MFP1 has
 CC features of a novel anchor protein that most likely connects chromatin

CC contains a putative wall/membrane anchoring domain. No homology was
 CC found, at the nucleotide and amino acid level, to any other submitted
 CC sequences in existing databases. The FBP binds horse fibrinogen, and
 CC protects against *S. equi* infection (which causes the disease strangles)
 CC in horses. The protein, or its truncated form, can be used for the
 CC preparation of a vaccine against strangles infection in horses.

XX Sequence 534 AA;

Query Match 18.0%; Score 93; DB 19; Length 534;

Best Local Similarity 29.9%; Pred. No. 0.45; Mismatches 51; Indels 30; Gaps 5;

Matches 41; Conservative 15; Mismatches 51; Indels 30; Gaps 5;

QY 2 ANNDVLRRLQ-----KGADQIIEYLKQVSL--LKEKAILQATLREKRLVE-- 51

DB 385 AASDAKVALEKEVEAAKAEVADLKQAKKEEELAEVKEKEALEAKIEELKTAHAEEL 444

QY 52 -----NAKLKEIEELKQEL--IOAEIONGVKQIAFPSTGPLHANSM----- 91

DB 445 SKLKEMLEKDHANADLOAEINRLKQELADRIKSLSGGRASQTNPGTTAKAGQLPSTG 504

QY 92 VSENVIOSTAVTVSSG 108

DB 505 ESANPFTIALTVIAG 521

RESULT 8

AAM98948

ID AAM98948 standard; Protein; 534 AA.

AC AAM98948;

DT 06-MAY-1999 (first entry)

XX Streptococcus equi M-protein.

XX Streptococcus equi; M-protein; Sem; immune response; horse; vaccine;

XX abscess; infection.

XX Streptococcus equi.

XX Key Location/Qualifiers

FT Peptide 1..36

FT Protein 37..534

FT /label= Sem

FT /note= "M-like protein"

FT Misc-difference 500..505

FT /note= "membrane anchor sequence"

XX MO9858945-A1.

XX 30-DEC-1998.

XX 23-JUN-1998; 98WO-US12962.

XX 23-JUN-1998; 98US-0103664.

XX 24-JUN-1997; 97US-0050577.

XX (ARTI/) ARTIUSHIN S.

XX (TIMO/) TIMONEY J F.

XX Artushin S, Timoney JF;

XX WPI; 1999-105605/09.

XX N-PSDB; AAX18839.

XX Compounds encoding the protective M-like Streptococcus equi protein

PT - useful for inducing *S. equi*-specific immunity and identifying *S.*

XX equi infection in horses

XX Claim 7, Page 6-7, 34pp; English.

CC The present sequence represents Sem, an M-like protein isolated from
 CC Streptococcus equi. A vaccine comprising Sem is useful for inducing
 CC *S. equi*-specific immunity in a horse. Sem, derived from a horse
 CC biological sample, can be detected in a PCR assay using primers given
 CC in AAX18840 and AAX18841, for identifying horses infected with *S. equi*.
 CC Compounds from the present invention enable quick assaying of *S. equi*.
 CC infection in horses, which is very important in the early stages of an
 CC outbreak. Vaccines comprising the compounds avoid the risk of reversion
 CC to virulence and occasional abscess development in vaccinated horses.

XX Sequence 534 AA;

Query Match 18.0%; Score 93; DB 20; Length 534;

Best Local Similarity 29.2%; Pred. No. 0.45; Mismatches 51; Indels 30; Gaps 4;

Matches 40; Conservative 16; Mismatches 51; Indels 30; Gaps 4;

QY 2 ANNDVLRRLQ-----KGADQIIEYLKQVSL--LKEKAILQATLREKRLVE-- 51

DB 385 AASDAKVALEKEVEAAKAEVADLKQAKKEEELAEVKEKEALEAKIEELKTAHAEEL 444

QY 52 -----NAKLKEIEELKQEL--IOAEIONGVKQIAFPSTGPLHANSM----- 91

DB 445 SKLKEMLEKDHANADLOAEINRLKQELADRIKSLSGGRASQTNPGTTAKAGQLPSTG 504

QY 92 VSENVIOSTAVTVSSG 108

DB 505 ESANPFTIALTVIAG 521

RESULT 9

AAM78507

ID AAM78507 standard; Protein; 1047 AA.

AC AAM78507;

DT 06-NOV-2001 (first entry)

XX Human protein SEQ ID NO 1169.

XX Human, cytokine; cell proliferation; cell differentiation; gene therapy;

XX vaccine; peptide therapy; stem cell growth factor; hematopoiesis;

XX tissue growth factor; immunomodulatory; cancer; leukaemia;

XX nervous system disorder; arthritis; inflammation.

XX Homo sapiens.

XX WO200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US04098.

XX 03-FEB-2000; 2000US-0496914.

XX 27-APR-2000; 2000US-0560875.

XX 20-JUN-2000; 2000US-0598075.

XX 19-JUL-2000; 2000US-0620325.

XX 01-SEP-2000; 2000US-0654936.

XX 15-SEP-2000; 2000US-0663561.

XX 20-OCT-2000; 2000US-0693125.

XX 30-NOV-2000; 2000US-0728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Dimaac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;

XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

XX Xue AJ, Yang Y, Wejthman T, Goodrich R;

XX WPI; 2001-476283/51.

XX N-PSDB; AAK51640.

XX Nucleic acids encoding polypeptides with cytokine-like activities,

PT useful in diagnosis and gene therapy -

PA (CURA-) CURAGEN CORP.
 PI Shinkets RA, Leach M;
 XX
 XX WPI; 2000-602362/57.
 DR N-PSDB; AAC75852.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorder,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 11; Page 2052-2053; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiproliferative; antipapillomatous; neurotropic; neuroprotective;
 CC osteoplastic; anticonvulsant; antidiabetic; immunosuppressive;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasodilator;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineoplastic;
 CC antihypertensive; antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 471 AA;
 Query Match 17.4%; Score 90; DB 21; Length 471;
 Best Local Similarity 27.0%; Pred. No. 0.78;
 Matches 30; Conservative 21; Mismatches 48; Indels 12; Gaps 3;
 QY 8 LKRLQGAADQI-----IEYLKQGVSLKEKAILQATLREKKLRVENAKLKEIE 60
 Db 129 VRELEQANDDLERAKRATYVSLFTLKLKQAIERNALFESLDEKESLIVSQRLKDEAR 188
 QY 61 ELKQELQATQNGVKGQIAPSGTPIANSMVSENVIQ---STAVTVSSG 108
 Db 189 DLROELAVREKQKQEVTRKSAFSSPTLCEKX--DSAVQASLSLPATPVGKG 237
 RESULT 12
 AAB96721 standard; Protein; 856 AA.
 AC AAB96721;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE Helicobacter pylori cellular proliferation protein #30.
 XX
 KM Antisense; prokaryotic cellular proliferation protein;
 KM antibiotic; antibacterial; drug design.
 XX
 OS Helicobacter pylori.
 XX
 PN WO200170955-A2.
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR

PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-263088P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 PI Haselbeck R, Ohlsen KL, Zykkind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 DR WPI; 2001-611495/70.
 DR N-PSDB; AAS53576.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 PS Example 3; Seq ID No 11310; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 856 AA;
 Query Match 17.2%; Score 89; DB 22; Length 856;
 Best Local Similarity 38.2%; Pred. No. 2.2;
 Matches 21; Conservative 14; Mismatches 20; Indels 0; Gaps 0;
 QY 11 LEOKGAADQIIEYLKQGVSLKEKAILQATLREKKLRVENAKLKEIE 65
 Db 428 MEKESNAKMOEILKELSDLKEEKIOLKQFENKEKVFKEISRLKMEKESLKE 482
 RESULT 13
 AAB96721 standard; Protein; 1177 AA.
 AC AAB96721;
 XX
 DT 29-OCT-2001 (first entry)
 XX
 DE Putative P. abyssi ATPase involved in DNA repair #3.
 XX
 KM Hyperthermophilic archaeon; hyperthermophilic protein.
 KM
 OS Pyrococcus abyssi.
 XX
 PN FR2792651-A1.
 PD 27-OCT-2000.
 XX
 PF 21-APR-1999; 99FR-0005034.
 XX
 PR 21-APR-1999; 99FR-0005034.
 PR
 PA (CNRS) CNRS CENT NAT RECH SCI.

PA (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
PI Querellou J, Weissenbach J, Saurin W, Heilig R;
XX WPI; 2001-126236/14.
XX
XX New nucleotide sequences isolated from *Pyrococcus abyssi* encode
PT proteins useful in industry -
XX
PS Claim 7, Pages 1483-1487, 1657pp; French.
XX
XX The present invention relates to the genomic sequence of *Pyrococcus*
CC *abyssi* (see AAP66431 and AAH41223-7) and P. *abyssi* proteins. P. *abyssi* is
CC a hyperthermophilic archaeon, which is isolated from deep-sea
CC hydrothermal vents. The present sequence is one such P. *abyssi* protein.
CC The proteins of the present invention have various potential industrial
CC uses, since the proteins are stable at very high temperatures, some up to
CC 110 degrees centigrade.
CC Note: This patent is in the same patent family as WO2000065062, which
CC contains additional sequences as shown in AAB99132-AAB99143,
CC AAH75903-AAH75920 and AAG66436.
XX
SQ Sequence 1177 AA;
XX
XX Query Match 17.2%; Score 89; DB 22; Length 1177;
XX Best Local Similarity 29.3%; Pred. No. 3.4; Indels 26; Gaps 3;
XX Matches 29; Conservative 20; Mismatches 24;
XX
QY 4 NDAVLKRLKQKGAADQIIIEYLK-----QOVSLEK-----ALLQA 40
DB 408 NEADIKRLKLEKERSRITLTKAKLPGRREVEKLEKLEKKAELSVENKTSISQR 467
QY 41 TLREKTLRVENAKLTK---EIEELKQELIQAEIQNVK 76
DB 468 RRKVEELKERTSELQKVSSELSERELIKAEQSEVR 506
XX
RESULT 14
ABG06864
ID ABG06864 standard; Protein; 472 AA.
XX
AC ABG06864;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #6855.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS71051.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX
PS Claim 20; SEQ ID No 37223; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 472 AA;
XX
XX Query Match 17.2%; Score 88.5; DB 22; Length 472;
XX Best Local Similarity 26.9%; Pred. No. 1.1; Indels 7; Gaps 2;
XX Matches 25; Conservative 29; Mismatches 32;
XX
QY 8 LKRLKQKGAADQIIIEYLKQVSLKKAIIQATLREKRLRVENAKLKEIELEKQELI 67
DB 383 LELDLVRYGQQAIEFLKEQISLAEKKML-----DYRLRLRAEQKQVEESLRKLL 436
QY 68 QAEIQ-NGYKQIAFSGTPIHANSVSENVIOS 99
DB 437 VAEENRLQAVEALCSSQHTMIESNDSEETIRT 469
XX
RESULT 15
AAO17717
ID AAO17717 standard; Protein; 667 AA.
XX
AC AAO17717;
XX
DT 08-AUG-2002 (first entry)
XX
DE O cuniculus endooligopeptidase A related protein.
XX
XX Human; protease associated protein; HPRAP; cancer; inflammation;
KM anti-HIV; antinaemic; antiarteriosclerotic; antiasthmatic; cytostatic;
KM antiinflammatory; hepatotropic; antidiabetic; nephrotropic; antiobese;
KM antichyroid; immunosuppressive; thyromimetic; virucide; dermatological;
KM neuroprotective; cardiant; osteopathic; antidiabetic; antiparasitic;
KM antipsoriatic; uropachic; ophthalmological; antirheumatic; tranquilliser;
KM vulnerary; antilucer; immune disorder; gene therapy; HPRAP-1.
XX
XX *Oryctolagus cuniculus*.
XX
XX WO9957274-A1.
XX
XX 11-NOV-1999.
XX
XX 28-APR-1999; 99WO-US09190.
XX
XX 01-MAY-1998; 98US-0071709.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Hillman JL, Tang YT, Lal P, Corley NC, Guegler KJ, Patterson C;
XX WPI; 2000-062147/05.
XX

XX Novel human protease associated proteins used for, e.g. the diagnosis
PT and prevention of cell proliferative and immune disorders
XX
PS
XX Disclosure; Page -; 75pp; English.

CC The present invention provides the protein and coding sequences of four
CC human protease associated proteins, designated HPRAP-1, HPRAP-2, HPRAP-3
CC and HPRAP-4. Human protease associated proteins are expressed in cancer
CC and immortalised cell lines and tissues associated with inflammation and
CC the immune response, and so appear to play a role in cell proliferative
CC and immune disorders. The sequences can be used to diagnose, treat or
CC prevent cell proliferative and immune disorders, including actinic
CC keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis,
CC hepatitis, mixed connective tissue disease, myelofibrosis, paroxysmal
CC nocturnal haemoglobinuria, polycythaemia vera, psoriasis, primary
CC thrombocythaemia, AIDS, Addison's disease, adult respiratory distress
CC syndrome, allergies, ankylosing spondylitis, amyloidosis, anaemia,
CC asthma, atherosclerosis, autoimmune haemolytic anaemia, autoimmune
CC thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's
CC disease, atopic dermatitis, dermatomyositis, diabetes mellitus,
CC emphysema, episodic lymphopenia with lymphocytotoxicity, erythroblastosis
CC fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis,
CC Goodpasture's syndrome, gout, Grave's disease, Hashimoto's thyroiditis,
CC hyperesoinophilia, irritable bowel syndrome, multiple sclerosis,
CC myasthenia gravis, myocardial or pericardial inflammation,
CC osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,
CC Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjogren's syndrome,
CC systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis,
CC thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome,
CC infections and trauma. The present sequence is the rabbit
CC endooligopeptidase A related protein, which was used in a homology
CC comparison in the exemplification of the invention.
CC Note: The present sequence was not obtainable from this Derwent basic but
CC was obtained from the Derwent family equivalent US6171790.

XX
SQ Sequence 667 AA;

Query Match 17.1%; Score 88; DB 21; Length 667;
Best Local Similarity 29.1%; Pred. No. 2;
Matches 32; Conservative 18; Mismatches 50; Indels 10; Gaps 3;

QY 2 ANNDVLRLEQKGAEDQIIIEYLRKQVSLKKAIIQATLRBEKKREVENAKLKEIE 61
Db 332 ANDD-----LERAKRATIVSLTETLTKNQAIERNAFLESELDKESILVSQRLDEARD 386
QY 62 LKQELIQAEIQNGVQKQIAFPSTPLHANSVSENVIO---STAVTVSSG 108
Db 387 LRQELAVRERQOEVTYRKSPSPFLDCEKM--DSAVQASLSLPATPVGKG 434

Search completed: January 2, 2004, 18:20:31
Job time : 33.1882 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:18:32 ; Search time 11.8588 seconds

(without alignments)
875.822 Million cell updates/sec

Title: US-09-930-169-2

Perfect score: 516

Sequence: 1 MANDAVIKRLKQKGAEDQ.....NSMSENVIOSTAVTTVSSG 108

Scoring table: BLOSUM62

Gapop: 10.0 , Gapext 0.5

Searched: 283308 segs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	511	99.0	310	2 B55053	endothelial monocy
2	395	76.6	310	2 A55053	endothelial monocy
3	101	19.6	270	2 G71061	endothelial monocy
4	96	18.6	163	2 D71169	endothelial monocy
5	94	18.2	213	2 B83972	endothelial monocy
6	93.5	18.1	662	2 A54078	endothelial monocy
7	93	18.0	1130	2 T34081	endothelial monocy
8	92	17.8	1169	2 A64505	endothelial monocy
9	91.5	17.7	507	2 S05542	endothelial monocy
10	90.5	17.5	162	2 A56254	endothelial monocy
11	89.5	17.3	160	2 S44868	endothelial monocy
12	89	17.2	143	2 H75062	endothelial monocy
13	89	17.2	856	2 H64552	endothelial monocy
14	89	17.2	1177	2 B75150	endothelial monocy
15	88	17.1	212	2 S42396	endothelial monocy
16	88	17.1	346	2 A55111	endothelial monocy
17	88	17.1	346	2 J67638	endothelial monocy
18	88	17.1	512	2 PC7064	endothelial monocy
19	87.5	17.0	1079	2 T38913	endothelial monocy
20	87.5	17.0	2116	2 A26653	endothelial monocy
21	87	16.9	234	2 I54074	endothelial monocy
22	86.5	16.8	389	2 B44972	endothelial monocy
23	86	16.7	955	2 S24348	endothelial monocy
24	86	16.7	1024	2 T30868	endothelial monocy
25	86	16.7	1256	2 T26101	endothelial monocy
26	86	16.7	1938	1 JX0178	endothelial monocy
27	85.5	16.6	638	2 S36723	endothelial monocy
28	85	16.5	348	2 C70417	endothelial monocy
29	85	16.5	369	2 I78877	endothelial monocy

30	85	16.5	508	2 A45477	CAMP response elem
31	85	16.5	752	2 A65089	DNA topoisomerase
32	85	16.5	752	2 G91116	DNA topoisomerase
33	85	16.5	752	2 G85961	DNA topoisomerase
34	85	16.5	752	2 T43788	DNA topoisomerase
35	84.5	16.4	375	2 T12251	DNA topoisomerase
36	84.5	16.4	662	2 D54078	hypothetical prote
37	84.5	16.4	662	2 D82814	hypothetical prote
38	84.5	16.4	1046	2 T42720	ATP-dependent Clp
39	84.5	16.4	1046	2 T42734	Cytoplasmic linker
40	84.5	16.4	1526	2 T41522	myosin ii - fisio
41	84.5	16.4	2288	2 T29959	hypothetical prote
42	84	16.3	400	2 E70318	hypothetical prote
43	84	16.3	752	2 A45582	DNA topoisomerase
44	84	16.3	1937	2 I38055	myosin heavy chain
45	83.5	16.2	394	2 T03998	ATG1 protein homol

ALIGNMENTS

```

RESULT 1
B55053
endothelial monocyte-activating protein II precursor - human
C:Species: Homo sapiens (man)
C:Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 09-Dec-2002
C:Accession: B55053
R:Kao, J.; Houck, K.; Fan, Y.; Haehnel, I.; Libutti, S.K.; Kayton, M.L.; Grikscheit, T.;
J., J.; Stern, D.M.
J. Biol. Chem. 269, 25106-25119, 1994
A:Title: Characterization of a novel tumor-derived cytokine. Endothelial-monocyte active
A:Reference number: A55053; MUID:95014290; PMID:7929199
A:Accession: B55053
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-310 <KAO>
A:Cross-references: GB:U0117
C:Superfamily: endothelial monocyte-activating polypeptide II precursor (pro-EMAP II)

Query Match          99.0%; Score 511; DB 2; Length 310;
Best Local Similarity 99.1%; Pred. No. 5.2e-31;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MANDAVIKRLKQKGAEDQIIETLKQVSLKSKKATIQATLRREKKLRVENAKLKEIE 60
DB 1 MANDAVIKRLKQKGAEDQIIETLKQVSLKSKKATIQATLRREKKLRVENAKLKEIE 60
QY 61 ELKQELIQAEIQNGVKQIAPSGTPLHANSWSENVIOSTAVTTVSSG 108
DB 61 ELKQELIQAEIQNGVKQIAPSGTPLHANSWSENVIOSTAVTTVSSG 108

RESULT 2
A55053
endothelial monocyte-activating protein II precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 09-Dec-2002
C:Accession: A55053; MUID:95014290; PMID:7929199
R:Kao, J.; Houck, K.; Fan, Y.; Haehnel, I.; Libutti, S.K.; Kayton, M.L.; Grikscheit, T.;
J., J.; Stern, D.M.
J. Biol. Chem. 269, 25106-25119, 1994
A:Title: Characterization of a novel tumor-derived cytokine. Endothelial-monocyte active
A:Reference number: A55053; MUID:95014290; PMID:7929199
A:Accession: A55053
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-310 <KAO>
A:Cross-references: GB:U0118; NID:g498911; PID:AAA62203.1; PID:g498912
R:Kao, J.; Ryan, J.; Bretz, G.; Chen, J.; Shen, H.; Fan, Y.G.; Godman, G.; Familletti, P
J. Biol. Chem. 267, 20239-20247, 1992
A:Title: Endothelial monocyte-activating polypeptide II. A novel tumor-derived polypeptide
A:Reference number: A44032; MUID:93015897; PMID:1400342
A:Accession: A44032

```

A:Molecule type: protein
A:Residues: 145-158, 'X', 160-164 <KA2>
A:Experimental source: methylcholanthrene A fibrosarcoma cells
A:Title: sequence extracted from NCBI backbone (NCBI:P115676)
C:Superfamily: endothelial monocyte-activating polypeptide II precursor (pro-EMAP II)

Query Match 76.6%; Score 395; DB 2; Length 310;
Best Local Similarity 77.6%; Pred. No. 2.1e-22;
Matches 83; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

QY 1 MANDAVLRLEKQKGAADIIIEYLKQVSLKEKAILQATLREKKLRVENAKKKEKE 60
DB 1 MATNDVLRLEKQKGAADIIIEYLKQVSLKEKAILQATLREKKLRVENAKKKEKE 60

QY 61 ELKQELIOAEIONGVQKQIAPPSGTPHANSVSENVISTPAVTSS 107
DB 61 ELKQELIOAEIONGVQKQIAPPSGTPHANSVSENVISTPAVTSS 107

RESULT 3
G71061
hypothetical protein PH189 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999
C/Accession: G71061

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kusuda, N.; Oguchi
DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:9834137; PMID:9679194

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Residues: 1-270 <KAW>
A:Cross-references: GB:AP000005; NID:G3236132; PIDN:BA30289.1; PID:dl031232; PID:G32576
A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH189

Query Match 19.6%; Score 101; DB 2; Length 270;
Best Local Similarity 30.0%; Pred. No. 1.2;
Matches 30; Conservative 22; Mismatches 32; Indels 16; Gaps 3;

QY 3 NNDVAVLRLEKQKGAADIIIEYLKQVSLKEKAILQATLREKKLRVENAKK 55
DB 86 NNDVAVLRLEKQKGAADIIIEYLKQVSLKEKAILQATLREKKLRVENAKK 55

QY 56 KKEIEELKQELIOAEIONGVQKQIAPPSGTPHANSVSENV 95
DB 144 KKEIEELKQELIOAEIONGVQKQIAPPSGTPHANSVSENV 95

RESULT 4
D71169
hypothetical protein PH0552 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C/Accession: D71169

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kusuda, N.; Oguchi
DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:9834137; PMID:9679194

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Residues: 1-163 <KAW>
A:Cross-references: GB:AP000002; NID:G3236129; PIDN:BA29641.1; PID:G3256958

A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH0552

Query Match 18.6%; Score 96; DB 2; Length 163;
Best Local Similarity 33.3%; Pred. No. 1.7;
Matches 24; Conservative 19; Mismatches 21; Indels 8; Gaps 2;

QY 8 LKRLKQKGAADIIIEYLKQVSLKEKAILQATLREKKLRVENAKKKEIEELKQEL 67
DB 83 LKRLKQKGAADIIIEYLKQVSLKEKAILQATLREKKLRVENAKKKEIEELKQEL 67

QY 68 OAEIONGVQKQI 79
DB 136 -OAEIONGVQKQI 146

RESULT 5
E83972
hypothetical protein BH2581 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C/Accession: E83972

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hiran
Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: AB3650; MUID:20512582; PMID:11058132

A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-213 <STO>
A:Cross-references: GB:AP001516; GB:BA000004; NID:gl0175192; PIDN:BA06300.1; GSPDB:GN001
A:Experimental source: strain C-125

C:Genetics:
A:Gene: BH2581

Query Match 18.2%; Score 94; DB 2; Length 213;
Best Local Similarity 36.8%; Pred. No. 3.2;
Matches 25; Conservative 16; Mismatches 15; Indels 12; Gaps 3;

QY 7 VLKRLKQKGAADIIIEYLKQVSLKEKAILQATLREKKLRVENAKKKEIEELKQEL 57
DB 115 VLKRLKQKGAADIIIEYLKQVSLKEKAILQATLREKKLRVENAKKKEIEELKQEL 57

QY 58 EIEELKQEL 65
DB 172 EIEELKQEL 179

RESULT 6
A54078
methyl-accepting chemotaxis protein mcpB - Bacillus subtilis

C:Species: Bacillus subtilis
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Jun-2000
C/Accession: A54078; H69655

R:Hanson, D.W.; Ordal, G.W.
J. Biol. Chem. 269, 14038-14046, 1994

A:Title: Cloning and characterization of genes encoding methyl-accepting chemotaxis prote
A:Reference number: A54078; MUID:94245722; PMID:8188684

A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-662 <HAN>
A:Cross-references: GB:U29189; NID:G459687

A:Note: the sequence in GenBank entry BACMCPBP, release 106.0, (PID:G459688) has 360-N
A:Reference number: A54078; MUID:94245722; PMID:8188684

A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-662 <HAN>
A:Cross-references: GB:U29189; NID:G459687

A:Note: the sequence in GenBank entry BACMCPBP, release 106.0, (PID:G459688) has 360-N
A:Reference number: A54078; MUID:94245722; PMID:8188684

Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A'Reference number: A69580; MUID:98044033; PMID:9384377
 A'Accession: H69555
 A'Status: preliminary; nucleic acid sequence not shown; translation not shown
 A'Molecule type: DNA
 A'Resides: 1-359, 'N', 361-449, 'G', 451-662 <KUN>
 A'CROSS-references: GB:299119; GB:299120; GB:AL009126; NID:g2635613; PIDN:CA15115.1; PI
 A'Experimental source: strain 168
 A'Comment: Strains in which this protein is inactivated are defective in chemotaxis towa
 C'Genetics:
 A'Gene: MCPB
 C'Superfamily: Probable methyl-accepting chemotaxis transducer
 C'Keywords: chemotaxis; signal transduction; transmembrane protein

Query Match 18.1%; Score 93.5; DB 2; Length 662;
 Best Local Similarity 29.8%; Pred. No. 12;
 Matches 34; Conservative 16; Mismatches 37; Indels 27; Gaps 4;

Oy 5 DAVLRLKQKGAEDQIIEYLK---QVSLKEKALIA-----TLREKTLR 49
 Db 464 EAVVAGLGKSKDITSLRVINGIADQTNLALNALPAAPAGSGSGSVAAERKLA 523

Oy 50 VENAFLKKEIELKQELIAEIQNGVKQIAFPSTPLHANSVSENIQSTAVT 103
 Db 524 VQSDASAEIKELQIEIV-ARID-----TSLMFREVNQEVQSGLVVT 565

RESULT 7
 T34081
 C'Species: *Caenorhabditis elegans*
 C'Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C'Accession: T34081
 R'Miller, N.
 A'Description: The sequence of *C. elegans* coamid C02F12.
 A'Reference number: Z21473
 A'Accession: T34081
 A'Status: preliminary; translated from GB/EMBL/DBJ
 A'Molecule type: DNA
 A'Resides: 1-1130 <ML>
 A'CROSS-references: EMBL:U41545; PIDN:AA83190.1; CESP:C02F12.7
 A'Genetics:
 A'Gene: C02F12.7
 A'Introns: 11/2; 118/1; 156/2; 239/2; 308/2; 398/1; 458/3; 591/1; 640/3; 756/2; 790/2; 8

Query Match 18.0%; Score 93; DB 2; Length 1130;
 Best Local Similarity 24.6%; Pred. No. 23;
 Matches 29; Conservative 30; Mismatches 33; Indels 26; Gaps 4;

Oy 2 ANNDVRLKQKGAEDQIIEYLKQVSLKEKALIA-----KLRYENAKL 55
 Db 691 SDKDLLEELSK---NKNIEHLKQIAQLNEKISTETKDSLEKTIQLEIDNSK 746

Oy 56 KKEIEEL-----KQELQAEIQNGVKQIAFPSTPLHANSVSENIQSTAVT 99
 Db 747 SDQIEKLRVNDMDQWGTIKDELVKXNE--IKTISAKTQALLSENTVSESTLAS 802

RESULT 8
 A64505
 C'Species: *Methanococcus jannaschii*
 C'Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 02-Jun-2000
 C'Accession: A64505
 R'Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.T.; Overbeek, R.; Kirkness, E.F.; Weissbrock, K.G.; Weirick, J.M.; Glodek, A.; Ison, J.D.; Sadow, P.M.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273, 1058-1073, 1996
 A'Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C

A>Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*.
 A'Reference number: A64300; MUID:96337999; PMID:8688087
 A'Accession: A64505
 A'Status: preliminary; nucleic acid sequence not shown; translation not shown
 A'Molecule type: DNA
 A'Resides: 1-1169 <BUL>
 A'CROSS-references: GB:U67604; GB:U77117; NID:g1592224; PID:g1500543; TIGR:MJ1643
 C'Genetics:
 A'Map position: FOR1623481-1626990
 C'Superfamily: chromosome segregation protein SMC1

Query Match 17.8%; Score 92; DB 2; Length 1169;
 Best Local Similarity 26.6%; Pred. No. 28;
 Matches 25; Conservative 24; Mismatches 25; Indels 20; Gaps 4;

Oy 3 NNDVRLKQKGAEDQIIEYLKQ---QVSLKEKALIA-----TLREKTLR 49
 Db 281 NN--IINELNEKNE--EVLHLKSIKELEIENDKVLDSINELKVEIEINKKE 336

Oy 59 IEELKQELIQ-----AEIQNGVKQIAF 80
 Db 337 IKETQKTIENRDSITKEQKIEIKKINLV 370

RESULT 9
 S05542
 C'Species: *Enterococcus faecium*
 C'Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 15-Oct-1999
 C'Accession: S05542
 R'Fuert, P.; Moesch, H.U.; Solioz, M.
 Nucleic Acids Res. 17, 6724, 1989
 A>Title: A protein of unusual composition from *Enterococcus faecium*.
 A'Reference number: S05542; MUID:89385998; PMID:2780297
 A'Accession: S05542
 A'Molecule type: DNA
 A'Resides: 1-507 <FUB>
 A'CROSS-references: GB:X16421; EMBL:M26048; NID:g43333; PIDN:CAA34442.1; PID:g43334
 A'Note: the authors translated the codon CGT for residues 221 and 223 as Lys

Query Match 17.7%; Score 91.5; DB 2; Length 507;
 Best Local Similarity 30.3%; Pred. No. 12;
 Matches 33; Conservative 28; Mismatches 37; Indels 11; Gaps 4;

Oy 5 DAVLRLKQKGAEDQIIEYLKQVSLKEKALIA-----KLRYENAKL 55
 Db 29 DKTIADIQNOQASQSQIEALKEGVSAINTKA--QDLTKQDTLRKSAQLKQIKDLOE 86

Oy 65 --ELIQAEIQNGVKQIAFPSTPLHANSVSENIQSTAVT 105
 Db 87 RIEKREATIQKARETOVKNTSSNYIDAVLNADSL-ADAVGRIOAMSTI 134

RESULT 10
 A56254
 C'Species: *Gallus gallus* (chicken)
 C'Date: 05-May-1995 #sequence_revision 05-May-1995 #text_change 21-Jul-2000
 C'Accession: A56254
 R'Kataoka, K.; Igatahshi, K.; Itoh, K.; Fujiwara, K.T.; Noda, M.; Yamamoto, M.; Nishizawa
 Mol. Cell. Biol. 15, 2180-2190, 1995
 A>Title: Small Mat proteins heterodimerize with Fos and may act as competitive repressors.
 A'Reference number: A56254; MUID:95198738; PMID:7891713
 A'Accession: A56254
 A'Status: preliminary; translated from GB/EMBL/DBJ
 A'Molecule type: DNA
 A'Resides: 1-162 <KAT>
 A'CROSS-references: GB:D28602; NID:g516725; PIDN:BA05939.1; PID:g516726
 A'Note: submitted to DDBJ, February 1994
 C'Genetics:
 A'Gene: matg
 A'Introns: 12/3

C/Superfamily: maf homology
C/Keywords: nucleus; transcription factor
F:25-114/Domain: maf homology <MAF>

Query Match 17.5%; Score 90.5; DB 2; Length 162;
Best Local Similarity 27.6%; Pred. No. 4.3;
Matches 29; Conservative 23; Mismatches 42; Indels 11; Gaps 3;

QY 9 KRLEOKGAEADQIIEYVKQVSLKEKAILQATLREKKLRVENAKLKKIEELKQELI 68
DB 57 RILKMGVYASCKVKTQCELEKQVLEQ---QVEKIASNASMKELDLRSK--Y 111
QY 69 AEIQNGVKQIAPSGTPLHA-----NSMVSENVIOSTAVTVSS 107
DB 112 EALQNFARTVARSPTVPVGRPLNSMGPVLPGRVATTSVITVKS 156

RESULT 11

844868
Kinesin heavy chain unc-116 - Caenorhabditis elegans

N/Alternate names: R05D3.7 protein

C/Species: Caenorhabditis elegans

C/Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 23-Mar-2001

C/Accession: S44868; A48264

R/Milson, R.

submitted to the EMBL Data Library, December 1992

A/Description: Sequence of the C. elegans cosmid R05D3.

A/Reference number: S44733

A/Accession: S44868

A/Molecule type: DNA

A/Residues: 1-843 <NTL>

A/Cross-references: EMBL:L07144; NID:G156412; PIDN:AAA28137.1; PID:G156420

R/Patel, N.; Thierri-Mieg, D.; Mancillas, J.R.

Proc. Natl. Acad. Sci. U.S.A. 90, 9181-9185, 1993

A/Title: Cloning by insertional mutagenesis of a cDNA encoding Caenorhabditis elegans ki

A/Reference number: A48264; MUID:94022338; PMID:8105472

A/Accession: A48264

A/Status: not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 1-814, 'X' <PAT>

A/Cross-references: GB:U19120; NID:G439589; PIDN:AAA28155.1; PID:G439590

A/Experimental source: mixed-stage hermaphrodite library

A/Note: sequence extracted from NCBI backbone (NCBIP:138249)

C/Genetics:

A/Intons: 13/3; 45/3; 99/3; 199/1; 275/3; 717/3

C/Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology

C/Keywords: ATP; nucleotide binding; P-loop

F:12-335/Domain: kinesin motor domain homology <KMOT>

F:88-95/Region: nucleotide-binding motif A (P-loop)

Query Match 17.3%; Score 89.5; DB 2; Length 843;

Best Local Similarity 36.0%; Pred. No. 30;

Matches 27; Conservative 22; Mismatches 17; Indels 9; Gaps 5;

QY 9 KRLEOKGAEADQIIEYVKQVSLKEKAILQATLREKKLRVENAKLKKIEELKQ 64
DB 440 QQDERDDDEQKXQSELEKQV-LIQERAL--GTRKREBELIRENNRFGQEAEDKQ 496
QY 65 ELIQNGVKQIAPSGTPLHA-----NSMVSENVIOSTAVTVSS 107
DB 497 E--GKEWMTALEIA 509

RESULT 12

H75062
flagella-related protein C PAB1381 - Pyrococcus abyssi (strain Orsay)

C/Species: Pyrococcus abyssi

C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C/Accession: H75062

R/anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A/Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru

A/Reference number: A75001

A/Accession: H75062
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-160 <KAW>
A/Cross-references: GB:AJ248287; GB:AL096836; NID:G5458657; PIDN:CA850397.1; PID:G545891C
A/Experimental source: strain Orsay
C/Genetics:
A/Genes: PAB1381

Query Match 17.2%; Score 89; DB 2; Length 160;
Best Local Similarity 30.6%; Pred. No. 5.4;
Matches 22; Conservative 21; Mismatches 21; Indels 8; Gaps 2;

QY 8 KRLEOKGAEADQIIEYVKQVSLKEKAILQATLREKKLRVENAKLKKIEELKQELI 67
DB 80 IERLDKTIQDMMLEKVEVSGEINPFEQ-----LSQSSLSFVQELAKQIEIRMEI- 132
QY 68 QAEIQNGVKQIAPSGTPLHA-----NSMVSENVIOSTAVTVSS 107
DB 133 -AQIKNDIKVLA 143

RESULT 13

H64552
endopeptidase Clp ATP-binding chain B - Helicobacter pylori (strain 26695)

N/Alternate names: ATP-dependent Clp proteinase regulatory chain

N/Contains: adenosinetriphosphatase (EC 3.6.1.3)

C/Species: Helicobacter pylori

C/Date: 09-Aug-1997 #sequence_revision 15-Aug-1997 #text_change 19-Jan-2001

C/Accession: H64552

R/Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalaj, H.G.; Glodek, A.; McKenney

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weldman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.

Nature 388, 539-547, 1997

A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser, C.

A/Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A/Reference number: A64520; MUID:97394467; PMID:9252185

A/Accession: H64552

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-856 <TOM>

A/Cross-references: GB:AE000545; GB:AE000511; NID:G2313349; PIDN:AA07330.1; PID:G231335C

C/Function:

A/Description: allows clp to hydrolyze polypeptides and proteins, probably by a chaperon

e activity; ATP hydrolysis is required for Clp hydrolysis of proteins but not of smaller

C/Superfamily: endopeptidase Clp ATP-binding chain

C/Keywords: ATP; duplication; hydrolase; molecular chaperone; nucleotide binding; P-loop

F:204-211/Region: nucleotide-binding motif A (P-loop)

F:271-276/Region: nucleotide-binding motif B

F:605-612/Region: nucleotide-binding motif A (P-loop)

F:673-678/Region: nucleotide-binding motif B

F:210/Binding site: ATP (lys) #status predicted

F:611/Binding site: ATP (lys) #status predicted

Query Match 17.2%; Score 89; DB 2; Length 856;
Best Local Similarity 38.2%; Pred. No. 34;
Matches 21; Conservative 14; Mismatches 20; Indels 0; Gaps 0;

QY 11 LRQKGAADQIIEYVKQVSLKEKAILQATLREKKLRVENAKLKKIEELKQELI 65
DB 428 MEKESNAKRMQBIKELSDLKEKIQLEAQFENEKEVKEISRLMEMESLKE 482

RESULT 14

B75150
chromosome segregation protein (smc1) PAB2109 - Pyrococcus abyssi (strain Orsay)

C/Species: Pyrococcus abyssi

C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C/Accession: B75150

R/anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A/Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru

A/Reference number: A75001

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: January 2, 2004, 18:17:12 ; Search time 8.47059 Seconds
(without alignments)
599.590 Million cell updates/sec

Title: US-09-930-169-2

Perfect score: 516
Sequence: 1 MANDAVIKLEKQKGAEDQ.....NSMSENVIOSTAVTVSSG 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	516	100.0	312	1	MCAL_HUMAN
2	425	82.4	359	1	MCAL_CRICK
3	395	76.6	310	1	MCAL_MOUSE
4	93.5	18.1	662	1	MCBP_BACU
5	93	18.0	1130	1	YL17_CABEL
6	92	17.8	1169	1	SMC_METUA
7	91.5	17.7	162	1	MARF_HUMAN
8	91.5	17.7	516	1	PS4_ENTFC
9	90.5	17.5	162	1	MARF_CHICK
10	90	17.4	956	1	KFSC_MOUSE
11	89.5	17.3	162	1	MARF_MOUSE
12	89.5	17.3	815	1	KINH_CABEL
13	89	17.2	856	1	CLPB_HELP
14	88	17.1	291	1	USF2_RAT
15	88	17.1	346	1	USF2_MOUSE
16	87.5	17.0	1079	1	IFP2_SCHPO
17	87.5	17.0	2116	1	MYS2_DICDI
18	87	16.9	346	1	USF2_HUMAN
19	86	16.7	1024	1	RIP3_MOUSE
20	86	16.7	1029	1	RIP3_RAT
21	86	16.7	1938	1	MYS2_CHICK
22	85.5	16.6	638	1	YAB1_YEAST
23	85.5	16.6	570	1	STIM_DROME
24	85	16.5	752	1	PARC_ECOLI
25	85	16.5	1941	1	MYH2_HUMAN
26	84.5	16.4	736	1	MYH7_BACU
27	84.5	16.4	1526	1	MYS2_SCHPO
28	84.5	16.4	752	1	PARC_SALTY
29	84	16.3	1937	1	MYH8_HUMAN
30	84	16.3	1939	1	MYH1_HUMAN
31	84	16.3	1934	1	MYH7_MESAU
32	83.5	16.2			
33	83.5	16.2			

34	83	16.1	171	1	ATPF_HELPY
35	83	16.1	452	1	TIG_CAUCR
36	83	16.1	724	1	HMWR_HUMAN
37	83	16.1	1013	1	AGOD_DROME
38	83	16.1	1093	1	TRF1_HUMAN
39	83	16.1	1690	1	C190_DROME
40	83	16.1	1939	1	MYH4_HUMAN
41	83	16.1	1940	1	MYH3_CHICK
42	82.5	16.0	790	1	MUS2_THETN
43	82.5	16.0	794	1	HMWR_MOUSE
44	82.5	16.0	1084	1	MYS2_RABIT
45	82.5	16.0	1935	1	MYH7_HUMAN

ALIGNMENTS

RESULT 1
MCAL_HUMAN STANDARD; PRT; 312 AA.
AC Q12904; Q96CQ9; 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Multisynthetase complex auxiliary component p43 [Contains:
DE Endothelial-monocyte activating polypeptide II (EMAP-II) (Small
DE inducible cytokine subfamily B member 1)].
CN SCYB1 OR EMAP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RX MEDLINE=95014290; PubMed=7929199;
RP KAO J., Houck K., Fan Y., Haehnel I., Libutti S.K., Kayton M.L.,
RA Grikscheit T., Chabot J., Nowyrod R., Greenberg S., Kuang W.J.,
RA Leung D.W., Hayward J.R., Kistel W., Heath M., Brett J., Stern D.M.,
RT "Characterization of a novel tumor-derived cytokine. Endothelial-
RT monocyte activating polypeptide II.",
RL J. Biol. Chem. 269:25106-25119 (1994).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Alekschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stjepanec M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshimiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosnak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Faley J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallie D.E.,
RA Schenck A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.",
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [1- SMILARITY: Contains 1 tRNA-binding domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>)

CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL: U0117; AA62202.1; -.
DR EMBL: BC014051; AA014051.1; -.
DR PDB; 1E72; 06-FEB-01.
DR PDB; 1E07; 06-SEP-01.
DR PDB; 1FLO; 07-FEB-01.
DR Genew; HGNC:10648; SCYE1.
DR MIM; 603605; -.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0005125; F:cytokine activity; TAS.
DR GO; GO:0000049; F:RNA binding activity; TAS.
DR GO; GO:0006418; P:amino acid activation; TAS.
DR GO; GO:0006935; P:chemotaxis; TAS.
DR GO; GO:0006954; P:inflammatory response; TAS.
DR InterPro; IPR002547; tRNA_bind.
DR Pfam; PF01588; tRNA_bind_1.
DR PIRSF; PIRSF005381; EMAP11; 1.
DR PROSITE; PS50886; TRBD; 1.
KM Protein biosynthesis; RNA-binding; tRNA-binding; cytokine;
KW 3D-structure.
FT PROSP 1 146
FT CHAIN 147 312
FT DOMAIN 151 252
FT CONFLICT 79 79 A -> P (IN REF. 2).
SQ SEQUENCE 312 AA; 34326 MW; 946310A0216F7587 CRC64;

Query Match 100.0%; Score 516; DB 1; Length 312;
Best Local Similarity 100.0%; Pred. No. 9e-31;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANNDAVLRKLEKGAADQIIIEYLKQOVSILKEKALIQATLREKKLRVENAKLKEIE 60
DB 1 MANNDAVLRKLEKGAADQIIIEYLKQOVSILKEKALIQATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIQAEIQNGVQKQIAFP--SGTPLHANSVSENVYQSTAVTAVTSSG 108
DB 61 ELKQELIQAEIQNGVQKQIAFP--SGTPLHANSVSENVYQSTAVTAVTSSG 108

RESULT 2
MCAL CRIGR STANDARD; PRT; 359 AA.
AC 054873;
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 28-FEB-2003 (Rel. 41; Last annotation update)
DE Multisynthetase complex auxiliary component p43 [Contains:
DE Endothelial-monocyte activating polypeptide II (EMAP-II) (Small
DE Inducible cytokine subfamily E member 1)].
GN SCYE1.
OS Cricketus gryllus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxId=10029;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98070438; Pubmed=9405472;
RA Quesillon S., Agou F., Robinson J.-C., Mirande M.;
RT "The p43 component of the mammalian multi-synthetase complex is likely
RT to be the precursor of the endothelial monocyte-activating polypeptide
RT II cytokine";
RL J. Biol. Chem. 272:32573-32579(1997).
CC -1- SUBUNIT: COMPONENT OF THE MULTISYNTHETASE COMPLEX WHICH IS
CC COMPOSED OF A BIFUNCTIONAL GLUTAMYL-PROLYL-TRNA SYNTHETASE, THE
CC MONOSPECIFIC ISOLUCYLYL, LEUCYLYL, GLUTAMINYLYL, METHIONYLYL, LYSYLYL,
CC ARGINYLYL, AND ASPARTYLYL-TRNA SYNTHETASES AS WELL AS THREE AUXILIARY
CC PROTEINS, P18, P48 AND P43.
CC -1- SIMILARITY: Contains 1 tRNA-binding domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF021800; AA0595207.1; -.
DR InterPro; IPR002547; tRNA_bind.
DR Pfam; PF01588; tRNA_bind_1.
DR PIRSF; PIRSF005381; EMAP11; 1.
DR PROSITE; PS50886; TRBD; 1.
KW Protein biosynthesis; RNA-binding; tRNA-binding; Cytokine.
FT DOMAIN 138 299
FT CHAIN 138 299
SQ SEQUENCE 359 AA; 39601 MW; 40868D1B65D72C23 CRC64;

Query Match 82.4%; Score 425; DB 1; Length 359;
Best Local Similarity 84.3%; Pred. No. 3.9e-24;
Matches 91; Conservative 8; Mismatches 7; Indels 2; Gaps 1;

QY 1 MANNDAVLRKLEKGAADQIIIEYLKQOVSILKEKALIQATLREKKLRVENAKLKEIE 60
DB 47 MANNDAVLRKLEKGAADQIIIEYLKQOVSILKEKALIQATLREKKLRVENAKLKEIE 106
QY 61 ELKQELIQAEIQNGVQKQIAFP--SGTPLHANSVSENVYQSTAVTAVTSSG 106
DB 107 ELKQELIQAEIQNGVQKQIAFPVQSDTPVQSSAVSTSVISQSTVSTIS 154

RESULT 3
MCAL MOUSE STANDARD; PRT; 310 AA.
AC P13130; O60659;
DT 01-JUL-1993 (Rel. 26; Last sequence update)
DT 01-JUL-1993 (Rel. 26; Last sequence update)
DT 15-SEP-2003 (Rel. 42; Last annotation update)
DE Multisynthetase complex auxiliary component p43 [Contains:
DE Endothelial-monocyte activating polypeptide II (EMAP-II) (Small
DE Inducible cytokine subfamily E member 1)].
GN SCYE1 OR EMAP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=95014290; Pubmed=7929199;
RA Kao J., Houck K., Fan Y., Haehnel I., Libutti S.K., Kayton M.L.,
RA Grikscheit T., Chabot J., Nowyrod R., Greenberg S., Kuang W.J.,
RA Leung D.W., Hayward J.R., Kistiel W., Heath M., Brett J., Stern D.M.;
RT "Characterization of a novel tumor-derived cytokine. Endothelial-
RT monocyte activating polypeptide II";
RL J. Biol. Chem. 269:25106-25119(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; Pubmed=12477932;
RA Strassberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Cantucci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.D., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultky S.W.,
RA Villalón D.K., Murry D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 145-164 FROM N.A.
RX MEDLINE=93015897; PubMed=1400342;
RA Kao J., Ryan J., Bretz G., Chen Y., Shen H., Fan Y.-G., Godman G.,
RA Fanillel P.C., Wang F., Pan Y.-C.E., Stern D., Claus M.;
RT "Endothelial monocyte-activating polypeptide II. A novel
RT tumor-derived polypeptide that activates host-response mechanisms.";
RN J. Biol. Chem. 267:20239-20247(1992).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=94193665; PubMed=7545917;
RA Kao J., Fan Y., Haehnel I., Bretz J., Greenberg S., Claus M.,
RA Kayton M., Houck K., Kiesel M., Seifeld R., Burnier J., Stern D.;
RT "A peptide derived from the amino terminus of endothelial-monocyte-
RT activating polypeptide II modulates mononuclear and polymorphonuclear
RT leukocyte functions, defines an apparently novel cellular interaction
RT site, and induces an acute inflammatory response.";
RN J. Biol. Chem. 269:9774-9782(1994).
CC -1- FUNCTION: ALTERS ENDOTHELIAL AND MONOCYTE FUNCTIONS, INDUCES THE
CC MIGRATION OF MONOCYTES AND GRANULOCYTES, AND INDUCES AN
CC INFLAMMATORY RESPONSE IN THE MOUSE FOOTPAD MODEL. EMAP II ELICITS
CC A PHLOGOGENIC RESPONSE AND, POTENTIALLY, AUGMENTS THE EFFECTS OF
CC THE OTHER TUMOR-DERIVED CYTOKINES.
CC -1- SUBUNIT: Monomer.
CC -1- SIMILARITY: Contains 1 trna-binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U10118; AAA62203.1; -;
DR EMBL; BC002054; AAA02054.1; -;
DR PIR; A55053; A55053.
DR MGI; 102774; Scyl1.
DR InterPro; IPR002547; trna_bind.
DR Pfam; PF01588; trna_bind.1.
DR PIRSF; PIRSF005381; EMAPII.1.
DR PROSITE; PS50886; TRBD.1.
DR Protein biosynthesis; RNA-binding; trna-binding; Cytokine.
FT PROPEP 1 144
FT CHAIN 145 310
FT DOMAIN 149 250
FT SEQUENCE 310 AA; 33997 MW; A2F8FF52A33D03A0 CRC64;
SQ
Query Match 76.6%; Score 395; DB 1; Length 310;
Best Local Similarity 77.6%; Pred. No. 5e-22;
Matches 83; Conservative 8; Mismatches 16; Indels 0; Gaps 0;
Qy 1 MANDAVYKRLKQGAADQIIIVYKQOVSLKEKALIIQATLRKKRVENAKIKKEIE 60
Db 1 MANDAVYKRLKQGAADQIIIVYKQOVSLKEKALIIQATLRKKRVENAKIKKEIE 60
Qy 61 ELKQELIAEIHNGEYVRVLSPLQNTASBSVGVSPVATAS 107
Db 61 ELKQELIAEIHNGEYVRVLSPLQNTASBSVGVSPVATAS 107
RESULT 4
MCPB_BACSU STANDARD; PRT; 662 AA.
AC P39215;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DR Methyl-accepting chemotaxis protein mcpB (H3).

GN MCPB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxId=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=168 / O11085;
RC MEDLINE=94245722; PubMed=8188684;
RA Hanton D.W., Ordal G.W.;
RT "Cloning and characterization of genes encoding methyl-accepting
RT chemotaxis proteins in Bacillus subtilis.";
RN J. Biol. Chem. 269:14038-14046(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RC MEDLINE=98044033; PubMed=9384377;
RA Kunst P., Ogatawa N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Berteiro M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot P., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Enlian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holstappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Kletter-Binchard M., Klein C.,
RA Kodayashl Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kunita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigne C.,
RA Medina N., Melillo R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudga B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Portwoll G., Prescott A.M.,
RA Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadleir Y.,
RA Sager T., Scallan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Takakoshi A., Tanaka T., Terpetra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambuit R., Wedler E., Wedler H., Weitenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumestein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -1- FUNCTION: RESPONSIBLE FOR DETECTING A SUBSET OF AMINO ACIDS THAT
CC INCLUDES ASPARAGINE, ASPARTATE, GLUTAMINE, AND HISTIDINE.
CC -1- FUNCTION: CHEMOTACTIC-SIGNAL TRANSDUCERS RESPOND TO CHANGES IN THE
CC CONCENTRATION OF ATTRACTANTS AND REPELLENTS IN THE ENVIRONMENT, AND
CC TRANSDUCE A SIGNAL FROM THE OUTSIDE TO THE INSIDE OF THE CELL, AND
CC FACILITATE SENSORY ADAPTATION THROUGH THE VARIATION OF THE LEVEL
CC OF METHYLATION. ALL AMINO ACIDS SERVE AS ATTRACTANTS IN
CC B. SUBTILIS, THEY APPEAR TO CAUSE AN INCREASE IN THE TURNOVER
CC METHYL GROUPS, LEADING TO METHYLATION OF AN UNIDENTIFIED ACCEPTOR,
CC WHILE REPELLENTS HAVE BEEN SHOWN TO CAUSE A DECREASE IN METHYL
CC GROUP TURNOVER. THE METHYL GROUPS ARE ADDED BY A METHYLTRANSFERASE
CC AND REMOVED BY A METHYLSTERASE.
CC -1- SIMILARITY: Contains 1 methyl-accepting transducer domain.
CC -1- SIMILARITY: Contains 1 hamp1 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L29189; AAA20554.1; -;
DR EMBL; Z99119; CAB15104.1; -;
DR EMBL; Z99120; CAB15115.1; -;


```

DR TIGR; MJI643; --
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR005289; GTP-binding_dom.
DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF02483; SMC_C.1.
DR Pfam; PF02463; SMC_N.1.
DR TIGR/Pfam; TIGR00650; MG442; 1.
KW Hypothetical protein; ATP-binding; Coiled coil; Complete proteome.
FT NP BIND 31 38 ATP (POTENTIAL).
FT DOMAIN 160 521 COILED COIL (POTENTIAL).
FT DOMAIN 673 1032 COILED COIL (POTENTIAL).
SQ SEQUENCE 1169 AA; 136634 MW; B63CE34E4C03F36 CRC64;

Query Match 17.8%; Score 92; DB 1; Length 1169;
Best Local Similarity 26.6%; Pred. No. 15;
Matches 25; Conservative 24; Mismatches 25; Indels 20; Gaps 4;

OY 3 NNDVAVLKELEKQAGADQIIIEYLKQ---QVSLKEKALQATTEKKLRVENAKLKE 58
DB 281 NN--IINELNKGNE--EVLHLKSIKELEVEIENDKVLQSSINELKKVEIENKKKE 336
OY 59 IEELKQELIQ-----AEIQNGVKQIAF 80
DB 337 IKETQKLIENRDSIIKEEQQIKELIEKIKMLNY 370

RESULT 7
MAGF_HUMAN STANDARD; PRT; 162 AA.
ID MAGF_HUMAN
DC O15525;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transcription factor MafG (V-maf musculoaponeurotic fibrosarcoma
DE oncogene homolog G) (hMaf).
GN MafG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97341189; PubMed=9195958;
RA Marini M.G., Chan K., Casula L., Kan Y.W., Cao A., Moir P.;
RT "hMaf, a small human transcription factor that heterodimerizes
RT specifically with Nrf1 and Nrf2."
RL J. Biol. Chem. 272:16490-16497(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Muscle;
RC MEDLINE=97309414; PubMed=9166829;
RA Blank V., Kim M.J., Andrews N.C.;
RT "Human MafG is a functional partner for p45 NF-E2 in activating
RT globin gene expression."
RL Blood 89:3925-3935(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=97432833; PubMed=9286713;
RA Blank V., Knoll J.H.M., Andrews N.C.;
RT "Molecular characterization and localization of the human MafG
RT gene."
RL Genomics 44:147-149(1997).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=97294601; PubMed=9150357;
RA Toki T., Itoh J., Kitazawa J., Arai K., Hatakeyama K., Akasaka J.,
RA Igarashi K., Nomura N., Yokoyama M., Yamamoto M., Ito E.;
RT "Human small Maf proteins form heterodimers with CNC family
RT transcription factors and recognize the NF-E2 motif."
RL Oncogene 14:1901-1910(1997).
RN [5]

```

```

RP SEQUENCE FROM N.A.
RA Ito E., Toki T.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stopleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loguélano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McMan P.U., McKernan K.J., Malek J.A., Gunnarac P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Wuzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA Paley J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Hyer R.M.,
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Scherich A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: SINCE THEY LACK A PUTATIVE TRANSCRIPTION DOMAIN, THE
CC SMALL MAFs BEHAVE AS TRANSCRIPTIONAL REPRESSORS WHEN THEY DIMERIZE.
CC AMONG THEMSELVES. HOWEVER, THEY SEEM TO SERVE AS TRANSCRIPTIONAL
CC ACTIVATORS BY DIMERIZING WITH OTHER (USUALLY LARGER) BASIC-ZIPPER
CC PROTEINS AND RECRUITING THEM TO SPECIFIC DNA-BINDING SITES. SMALL
CC MAF PROTEINS HETERODIMERIZE WITH POU AND MAY ACT AS COMPETITIVE
CC REPRESSORS OF THE NF-E2 TRANSCRIPTION FACTOR. TRANSCRIPTION
CC FACTOR, COMPONENT OF ERYTHROID-SPECIFIC TRANSCRIPTION FACTOR NF-
CC E2. ACTIVATES GLOBIN GENE EXPRESSION WHEN ASSOCIATED WITH NF-E2.
CC -1- SUBUNIT: HOMO- OR HETERODIMER. IT CAN ALSO FORM HIGH AFFINITY
CC HETERODIMERS WITH MEMBERS OF THE NFE2/CNC-BZIP FAMILY: NRF1,
CC NFE2 AND P45-NFE2
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN SKELETAL MUSCLE. ALSO
CC EXPRESSED IN HEART AND BRAIN.
CC -1- SIMILARITY: Belongs to the bZIP family. Maf subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; Y11514; CAAT2284.1; --
DR EMBL; U84249; AAC51737.1; --
DR EMBL; AF059195; AAC14427.1; --
DR EMBL; BC012327; AAH12327.1; --
DR TRANSFAC; T04870; --
DR Genew; HGNC:6781; MAFG.
DR MIM; 602020; --
DR InterPro; IPR004827; TF bZIP.
DR InterPro; IPR004826; TF Maf.
DR Pfam; PF01131; bZIP_Maf_1.
DR SMART; SM00338; BZIP_1.
DR PROSITE; PS50217; BZIP_1.
DR PROSITE; PS00036; BZIP_BASIC; FALSE NEG.
KW Transcription regulation; DNA-binding; Repressor; Nuclear protein.
FT DNA BIND 53 83 BASIC MOTIF.
FT DOMAIN 86 114 LEUCINE-ZIPPER.
SQ SEQUENCE 162 AA; 17849 MW; E49F1FBA230F8D0 CRC64;

Query Match 17.7%; Score 91.5; DB 1; Length 162;
Best Local Similarity 30.1%; Pred. No. 2.2;
Matches 31; Conservative 21; Mismatches 44; Indels 7; Gaps 3;

```

QY 9 KRLEOKGAEADQIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIEEL--KOEL 66
 DB 57 RLLKNGVYASCVKRVTKQKELEKQAKELQ---QEVKGLASNSMKLELDALRSKYEA 113
 QY 67 IQAEIONGVKQIAFPSCGTPPLHA--NSMVSSENVIOSTRAVTVSS 107
 DB 114 LQTFARTVARSVPVAPARGPLAAGLPGVAPKVAATSVITTVKS 156
 RESULT 8
 P54_ENTFC STANDARD; PRT; 516 AA.
 ID_P54_ENTFC STANDARD; PRT; 516 AA.
 AC P13692;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE P54 protein precursor.
 OS Enterococcus faecium (Streptococcus faecium).
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
 NCBI_TaxId=1352;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=69385998; PubMed=2780297;
 RA Feuer P., Moesch H.-U., Solioz M.;
 RT "A protein of unusual composition from Enterococcus faecium";
 RL Nucleic Acids Res. 17:6724-6724(1989).
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C40.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X16421; CAA34442.1; ALT_INT.
 DR PIR: S05542; S05542.
 DR MEROPS: C40.UPW.
 DR InterPro: IPR000064; NLP_C_P60.
 DR Pfam: PF00877; NLP_C_P60; 1.
 KM Signal; Cell wall.
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 516 P54 PROTEIN.
 FT ACT_SITE 429 429 POTENTIAL.
 SQ SEQUENCE 516 AA; 54596 MW; 402ECNA439846D26 CRC64;
 Query Match 17.7%; Score 91.5; DB 1; Length 516;
 Best Local Similarity 30.3%; Pred. No. 7.2;
 Matches 33; Conservative 28; Mismatches 37; Indels 11; Gaps 4;
 QY 5 DAVLKRLQKGAADQIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIEELKQ 64
 DB 38 DKRIADLQNGQASQISQIEALKEGVSAINTKA--QDLTKRQDILKRESAQDKQEIOLQ 95
 QY 65 --ELIQAELIONGVKQIAFPSCGTP-----LHNSMVSSENVIOSTRAVTV 105
 DB 96 RIEKREAFIQKQARETOVKNTSSNYIDAVLADSL-ADAVGRIQAMSTI 143
 RESULT 9
 MAFG_CHICK STANDARD; PRT; 162 AA.
 ID_MAFG_CHICK STANDARD; PRT; 162 AA.
 AC Q90889;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Transcription factor Mafg (V-maf musculoaponeurotic fibroblastoma
 DE oncogene homolog G).
 GN MAFG.
 OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NCBI_TaxId=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95198738; PubMed=7891713;
 RA Kataoka K., Igatahshi K., Itoh K., Fujiwara K.T., Noda M.,
 RA Yamamoto M., Nishizawa M.;
 RT "Small Maf proteins heterodimerize with Fos and may act as
 RT competitive repressors of the NF-E2 transcription factor";
 RL Mol. Cell. Biol. 15:2180-2190(1995).
 CC -1- FUNCTION: SINCE THEY LACK A PUTATIVE TRANSCRIPTION DOMAIN, THE
 CC SMALL MAFs BEHAVE AS TRANSCRIPTIONAL REPRESSORS WHEN THEY DIMERIZE
 CC AMONG THEMSELVES. HOWEVER, THEY SEEM TO SERVE AS TRANSCRIPTIONAL
 CC ACTIVATORS BY DIMERIZING WITH OTHER (USUALLY LARGER) BASIC-ZIPPER
 CC PROTEINS AND RECRUITING THEM TO SPECIFIC DNA-BINDING SITES. SMALL
 CC MAF PROTEINS HETERODIMERIZE WITH FOS AND MAY ACT AS COMPETITIVE
 CC REPRESSORS OF THE NF-E2 TRANSCRIPTION FACTOR. TRANSCRIPTION
 CC FACTOR, COMPONENT OF ERYTHROID-SPECIFIC TRANSCRIPTION FACTOR NF-
 CC E2.
 CC -1- SUBUNIT: HOMO- OR HETERODIMER.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: Belongs to the bZIP family. Maf subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D28602; BA05939.1; -
 DR PIR: I50378; A56254.
 DR TRANSFAC: T01437; -
 DR InterPro: IPR004827; TF_bZIP.
 DR InterPro: IPR004826; TF_Maf.
 DR Pfam: PF03131; bZIP_Maf.1.
 DR SMART: SM00338; BRLZ; 1.
 DR PROSITE: PS00217; BZIP; 1.
 DR PROSITE: PS00036; BZIP_BASIC; FALSE_NEG.
 DR Transcription regulation; DNA-binding; Repressor; Nuclear protein.
 KW DNA BIND 53 83 BASIC MOTIF.
 FT DOMAIN 86 114 LEUCINE-ZIPPER.
 SQ SEQUENCE 162 AA; 18077 MW; C7B0FCD188060596C CRC64;
 Query Match 17.5%; Score 90.5; DB 1; Length 162;
 Best Local Similarity 27.6%; Pred. No. 2.7;
 Matches 29; Conservative 23; Mismatches 42; Indels 11; Gaps 3;
 QY 9 KRLEOKGAEADQIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIEELKQELQ 68
 DB 57 RLLKNGVYASCVKRVTKQKELEKQAKELQ---QEVKGLASNSMKLELDALRSK--Y 111
 QY 69 ARIQNGVGVQIAFPSCGTPPLHA-----NSMVSSENVIOSTRAVTVSS 107
 DB 112 EALQNPARTVARSVPVAPARGPLTSMGVLPGKVAATSVITTVKS 156
 RESULT 10
 KFSC_MOUSE STANDARD; PRT; 956 AA.
 ID_KFSC_MOUSE STANDARD; PRT; 956 AA.
 AC P28738; Q922F8;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Kinesin heavy chain isoform 5C (Kinesin heavy chain neuron-specific
 DE 2).
 GN KFSC OR NKGHC2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.


```

OX NCB1_Taxid=10090;
RN [1]
RP PRELIMINARY SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Brain;
RA Kato K.;
RT "A collection of cDNA clones with specific expression patterns in
   mouse brain."
RL Eur. J. Neurosci. 2:704-711(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99000642; PubMed=9782088;
RA Xia C., Rahman A., Yang Z., Goldstein L.S.B.;
RT "Chromosomal localization reveals three kinesin heavy chain genes in
   mouse."
RL Genomics 52:209-213(1998).
CC -1- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING
CC PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT.
CC -1- SUBUNIT: OLIGOMER COMPOSED OF TWO HEAVY CHAINS AND TWO LIGHT
CC CHAINS.
CC -1- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-
CC TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF
CC KINESIN (IT HYDROLYZES ATP AND BINDS MICROTUBULE), A CENTRAL
CC ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN
CC DIMERIZATION, AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH
CC INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),
CC VESICLES AND MEMBRANOUS ORGANELLES.
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC SUBFAMILY.
CC -1- CAUTION: REF.1 SEQUENCE SEEMS TO HAVE BEEN A HYBRID OF A KIF5A
CC AND A KIF5C SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X61435; CAA43677.1; -.
DR EMBL; AF067180; AAC79804.1; -.
DR HSSP; P56536; 2KIN.
DR MGD; MGI:1098269; Klf5c.
DR GO; GO:0005515; F:protein binding activity; IPI.
DR GO; GO:0008045; P:motor axon guidance; IMP.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS00667; KINESIN_MOTOR_DOMAIN2; 1.
DR Motor protein; Microtubules; ATP-binding; Coiled coil.
FT DOMAIN 1 386 KINESIN_MOTOR (BY SIMILARITY).
FT DOMAIN 406 923 GLOBULAR.
FT DOMAIN 859 956 MICROTUBULE-BINDING.
FT DOMAIN 174 315 COILED COIL.
FT NP_BIND 86 93 ATP (BY SIMILARITY).
SQ SEQUENCE 956 AA; 109240 MW; D5A8C701A2911AB2 CRC64;
Query Match 17.4%; Score 90; DB 1; Length 956;
Best Local Similarity 26.1%; Pred. No. 17;
Matches 29; Conservative 25; Mismatches 43; Indels 14; Gaps 4;
OY 6 AVLRLEQKGAAD--QIIEYLKQVSLLEKAILQATLEBEK-----LRYENAKL 55
Db 422 SLVRLDQKDEINQOSLAELKQK--MLDDELASTRDYKIQEQLFRLEQENBA 479
OY KKEIEELQGLQAEIQNGVQIAPSPGPHANSMVENYQSAVTVTS 106
Db 480 KDEVEVQLQAELEAVNTQKQSEVEDKT--RANQLTDELAKTKTTTLTT 528

```

```

RESULT 11
ID KIF5C HUMAN STANDARD; PRT; 957 AA.
AC 060282; 095079;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kinesin heavy chain isoform 5C (Kinesin heavy chain neuron-specific
   2).
DN KIF5C OR NKHC2 OR KIAA0531.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCB1_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 5:31-39(1998).
RN [2]
RP SEQUENCE OF 355-585 FROM N.A.
RA Engelender S., Sharp A.H., Colomer V., Tokito M.K., Lanahan A.,
RA Wolley P., Holzbaur E.L.F., Ross C.A.;
RT "Huntingtin associated protein 1 (HAP1) interacts with the p150glued
RT subunit of dynein."
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING
CC PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT.
CC -1- SUBUNIT: OLIGOMER COMPOSED OF TWO HEAVY CHAINS AND TWO LIGHT
CC CHAINS.
CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN BRAIN, PROSTATE, AND
CC TESTIS, AND MODERATE EXPRESSION IN KIDNEY, SMALL INTESTINE, AND
CC OVARY.
CC -1- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-
CC TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF
CC KINESIN (IT HYDROLYZES ATP AND BINDS MICROTUBULE), A CENTRAL
CC ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN
CC DIMERIZATION, AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH
CC INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),
CC VESICLES AND MEMBRANOUS ORGANELLES.
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB011103; BAA25457.1; -.
DR EMBL; AF010146; AAD01436.1; -.
DR HSSP; P56536; 2KIN.
DR Genew; HGNC:6325; KIF5C.
DR MIM; 604593; -.
DR GO; GO:0005871; C:kinesin complex; TAS.
DR GO; GO:0003777; F:microtubule motor activity; TAS.
DR GO; GO:0006996; P:organelle organization and biogenesis; TAS.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS00667; KINESIN_MOTOR_DOMAIN2; 1.
DR Motor protein; Microtubules; ATP-binding; Coiled coil.
FT DOMAIN 1 386 KINESIN_MOTOR (BY SIMILARITY).
FT DOMAIN 406 923 COILED COIL.

```

```

FT DOMAIN 859 956 GLOBULAR.
FT DOMAIN 174 315 MICROTUBULE-BINDING.
FT NP BIND 86 93 ATP (BY SIMILARITY).
FT CONFLICT 355 360 TLNVI -> STNASH (IN REF. 2).
FT CONFLICT 583 585 EFT -> DRY (IN REF. 2).
SQ SEQUENCE 957 AA; 109494 MW; A9F25B1C994322A CRC64;

Query Match 17.4%; Score 90; DB 1; Length 957;
Best Local Similarity 26.1%; Pred. No. 17;
Matches 29; Conservative 25; Mismatches 43; Indels 14; Gaps 4;

QY 6 AVLKRLQKGAAD---QIIEYLKQVSLKEKAILQATLREKK-----LRVNAKL 55
DB 423 SLYRQDDKDEINQSLAEKLCQ--MDDQDELASTRDYEKIQEELRLQIENEA 480
QY 56 KKEIEELKQELIQAEIQNGVKQIAFPGGFLHANSWSENVISTAVTVSS 106
DB 481 KDEVKEVLQAEELAVNYDQKSGQVEDKT--RANEQLTDELAQRTTLTIT 529

RESULT 12
MARG_MOUSE STANDARD; PRT; 162 AA.
ID MARG_MOUSE
AC 054790;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transcription factor Mafg (V-maf musculoaponeurotic fibrosarcoma
DE oncogene homolog G).
GN MAFG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=129/SvJ;
RX MEDLINE=98344009; PubMed=9679061;
RA Shavit J.A., Mochizashi H., Onodera K., Akasaka J., Yamamoto M.,
RA Engel J.D.;
RT "Impaired megakaryopoiesis and behavioral defects in mafg-null mutant
RT mice.";
RL Genes Dev. 12:2164-2174(1998).
RN [2]
RP STRUCTURE BY NMR OF 24-64.
RX MEDLINE=21912421; PubMed=11875518;
RA Kusunoki H., Mochizashi H., Katsunaka F., Mochizashi A., Yamamoto M.,
RA Tanaka T.;
RT "Solution structure of the DNA-binding domain of Mafg.";
RL Nat. Struct. Biol. 9:252-256(2002).
CC -1- FUNCTION: SINCE THEY LACK A PUTATIVE TRANSCRIPTION DOMAIN, THE
CC SMALL MAPS BEHAVE AS TRANSCRIPTIONAL REPRESSORS WHEN THEY DIMERIZE
CC AMONG THEMSELVES. HOWEVER, THEY SEEM TO SERVE AS TRANSCRIPTIONAL
CC ACTIVATORS BY DIMERIZING WITH OTHER (USUALLY LARGER) BASIC-ZIPPER
CC PROTEINS AND RECRUITING THEM TO SPECIFIC DNA-BINDING SITES. SMALL
CC MAP PROTEINS HETERODIMERIZE WITH FOS AND MAY ACT AS COMPETITIVE
CC REPRESSORS OF THE NF-E2 TRANSCRIPTION FACTOR. TRANSCRIPTION
CC FACTOR, COMPONENT OF ERYTHROID-SPECIFIC TRANSCRIPTION FACTOR NF-
CC E2. ACTIVATES GLOBIN GENE EXPRESSION WHEN ASSOCIATED WITH NF-E2
CC (BY SIMILARITY).
CC -1- SUBUNIT: HOMO-OR HETERODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: Belongs to the bzip family. Maf subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC
CC EMBL; AB009693; BAA24028.1; -

```

```

DR PDB; 1KIV; 18-DEC-02.
DR MGD; MGI:96911; Mafg.
DR GO; GO:0030534; P:regul. behavior; IMP.
DR GO; GO:0042127; P:regulation of cell proliferation; IMP.
DR InterPro; IPR004827; TF_bZIP.
DR InterPro; IPR004826; TF_Maf.
DR Pfam; PF03131; bZIP_Maf; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00217; bZIP; 1.
DR PROSITE; PS00036; bZIP_BASIC; FALSE NEG.
KM Transcription regulation; DNA-binding; Repressor; Nuclear protein;
FT DNA BIND 53 83 BASIC MOTIF.
FT DOMAIN 86 114 LEUCINE-ZIPPER.
SQ SEQUENCE 162 AA; 17876 MW; C7FF19614EB95C7D CRC64;

Query Match 17.3%; Score 89.5; DB 1; Length 162;
Best Local Similarity 30.1%; Pred. No. 3.1;
Matches 31; Conservative 21; Mismatches 44; Indels 7; Gaps 3;

QY 9 KRLQKGAADQIIEYLKQVSLKEKAILQATLREKKLRVNAKLKKEIEEL--KOEL 66
DB 57 RLKRGVYASGVKRVYQKELEKQKALQ--QEVETLASENSMKLELALRSKYEA 113
QY 67 IQAEIQNGVKQIAFPGGFLHA--NSWSENVISTAVTVSS 107
DB 114 LQNFATVARSVPVAPRGFLAAGLPGVGVKVAATSVITVKS 156

RESULT 13
KINH_CAEEL STANDARD; PRT; 815 AA.
ID KINH_CAEEL
AC P34540;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Kinesin heavy chain.
GN UNC-116 OR R05D3.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=Bristol N2;
RX MEDLINE=94022338; PubMed=8105472;
RA Patel N., Thierry-Mieg D., Mancillas J.R.;
RT "Cloning by insertional mutagenesis of a cDNA encoding Caenorhabditis
RT elegans kinesin heavy chain.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:9181-9185(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Siddiqui S.S., Ali M.Y., Khan M.A.;
RT Submitted (Aug-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RC SEQUENCE FROM N.A.
RP STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favejlo A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kerhaw J., Kirsten J., Lalister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Koopra A., Saunders D., Shownkeen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstein L., Wilkinson-Sproat J.,
RA Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [4]

```

RP REVISIONS.
RA Waterston R.;
RL Submitted (Mar-2001) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING
CC PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT.
CC -1- SUBUNIT: OLIGOMER COMPOSED OF TWO HEAVY CHAINS AND TWO LIGHT
CC CHAINS.
CC -1- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-
CC TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF
CC KINESIN (IT HYDROLYSES ATP AND BINDS MICROTUBULE), A CENTRAL
CC ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN
CC DIMERIZATION, AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH
CC INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),
CC VESICLES AND MEMBRANOUS ORGANELLES.
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; L19120; AAA28155.1; -
CC EMBL; AB017163; BAA32594.1; -
CC EMBL; L07144; AAK21446.1; -
CC PIR; S44868; S44868.
CC HSSP; P33176; 1B62.
CC WormPep; R05D3.7; CE25070.
CC InterPro; IPR001752; kinesin_motor.
CC Pfam; PF00225; kinesin_1.
CC PRINTS; PR00380; KINESINHEAVY.
CC SMART; SM00129; KISC; 1.
CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
CC PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
CC Motor protein; Microtubules; ATP-binding; Coiled coil.
CC DOMAIN 1 259 KINESIN-MOTOR.
FT DOMAIN 335 374 COILED COIL (POTENTIAL).
FT DOMAIN 422 554 COILED COIL (POTENTIAL).
FT NP BIND 695 785 COILED COIL (POTENTIAL).
FT NP BIND 88 95 ATP (BY SIMILARITY).
SQ SEQUENCE 815 AA; 91893 MW; 1B32718C3A7C254C CRC64;

Query Match 17.3%; Score 89.5; DB 1; Length 815;
Best Local Similarity 36.0%; Pred. No. 16;
Matches 27; Conservative 22; Mismatches 17; Indels 9; Gaps 5;

QY 9 KRLEQKGAEDQI--IEVLKQVSLTKKAILQATLR-EKKLVENAKLKEIEELKQ 64
DB 440 QQLDEKDEIKVSELEKLRQV-LLOBEAL--GIMENBELIRENNRFOKEADKQ 496
QY 65 ELIQAEIONGVKQIA 79
DB 497 E--GKEMMTALAEIA 509

RESULT 14
CLPB_HELPY STANDARD; PRT; 856 AA.
AC P71404;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE CLPB protein.
GN CLPB OR HP0264.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN (1)
RP SEQUENCE FROM N.A.

RC STRAIN=NCTC 11637;
RA Allan E.;
RL Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEBLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kervatage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodge R., Khakhria G., Glodex A.,
RA McInerney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Feldman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547(1997).
CC -1- SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR
CC ATP-DEPENDENT PROTEASE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CLPB/CLPB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; Y08238; CAA69406.1; -
CC EMBL; AE000545; AAD07330.1; -
CC PIR; H64552; H64552.
CC TIGR; HP0264; -
CC InterPro; IPR003593; AAA ATPase.
CC InterPro; IPR003959; AAA ATPase cent.
CC InterPro; IPR004176; Chaprinin_clpA/B.
CC Pfam; PF00004; AAA; 2.
CC Pfam; PF02861; CLP_N; 2.
CC PRINTS; PR00300; CLPPTOTASEA.
CC SMART; SM00382; AAA; 2.
CC PROSITE; PS00870; CLPB 1; 1.
CC PROSITE; PS00871; CLPB 2; 1.
KM Chapterone; ATP-binding; Repeat; Complete proteome.
FT DOMAIN 159 406 1.
FT NP BIND 531 722 11.
FT NP BIND 204 211 ATP (POTENTIAL).
FT NP BIND 605 612 ATP (POTENTIAL).
FT CONFLICT 14 14 T -> A (IN REF. 1).
FT CONFLICT 33 33 M -> L (IN REF. 1).
FT CONFLICT 59 59 Q -> E (IN REF. 1).
FT CONFLICT 62 62 R -> K (IN REF. 1).
FT CONFLICT 79 79 S -> N (IN REF. 1).
FT CONFLICT 103 103 R -> T (IN REF. 1).
FT CONFLICT 119 119 G -> S (IN REF. 1).
FT CONFLICT 131 131 A -> T (IN REF. 1).
FT CONFLICT 145 145 R -> A (IN REF. 1).
FT CONFLICT 149 149 D -> G (IN REF. 1).
FT CONFLICT 222 222 M -> V (IN REF. 1).
FT CONFLICT 435 435 A -> H (IN REF. 1).
FT CONFLICT 465 465 V -> A (IN REF. 1).
FT CONFLICT 643 643 MS -> IT (IN REF. 1).
FT CONFLICT 649 649 A -> P (IN REF. 1).
FT CONFLICT 734 734 E -> D (IN REF. 1).
FT CONFLICT 765 765 D -> G (IN REF. 1).
FT CONFLICT 836 836 V -> I (IN REF. 1).
FT CONFLICT 840 840 D -> G (IN REF. 1).
SQ SEQUENCE 856 AA; 96683 MW; E3902C7D98946AD CRC64;

Query Match 17.2%; Score 89; DB 1; Length 856;
Best Local Similarity 38.2%; Pred. No. 18;

Matches 21; Conservative 14; Mismatches 20; Indels 0; Gaps 0;
 QY 11 LEQGAADQIIIEYIKQOVSILKEKALIQATLREKKLRVENAKLKEIEELKOE 65
 Db 428 MEKESNAKMOILKELSLKKEKIQLEKQFENEKEVEFKEISRLKEMESLKEKE 482

RESULT 15

USF2 RAT STANDARD; PRT; 291 AA.
 AC 063665;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Upstream stimulatory factor 2 (Upstream transcription factor 2)
 DE (Major late transcription factor 2) (Fragment).
 GN USF2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 ON NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RX MEDLINE=96139468; PubMed=8576131;
 RA Viollet B., Lefrançois-Martinez A.-M., Henrion A., Kahn A.,
 RA Raymondjean M., Martinez A.;
 RT "Immunological characterization and transacting properties of
 upstream stimulatory factor isoforms."
 RL J. Biol. Chem. 271:1405-1415(1996).
 CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO A SYMMETRICAL DNA
 CC SEQUENCE (E-BOXES) (5'-CAGTG-3') THAT IS FOUND IN A VARIETY OF
 CC VIRAL AND CELLULAR PROMOTERS.
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
 CC BHLH PROTEIN. BINDS DNA AS AN HOMODIMER OR A HETERODIMER
 CC (USF1/USF2).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=1;
 CC Comment=At least 2 isoforms are produced;
 CC Name=1;
 CC IsoId=Q63665-1; Sequence=Displayed;
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X90823; CA62338.1; -.
 CC DR HSSP; P22415; 1AN4.
 CC DR TRANSPAC; T02115; -.
 CC DR InterPro; IPR001092; HLH_basic.
 CC DR Pfam; PF00010; HLH_1.
 CC DR SMART; SM00353; HLH_1.
 CC DR PROSITE; PS00038; HLH_1; 1.
 CC DR PROSITE; PS50888; HLH_2; 1.
 CC KM DNA-binding; Nuclear protein; Transcription regulation;
 CC Alternative splicing.
 CC FT NON_TER 1
 CC FT DNA_BIND 180 193 BASIC DOMAIN.
 CC FT DOMAIN 194 236 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
 CC FT DOMAIN 252 273 LEUCINE-ZIPPER (POTENTIAL).
 CC SQ SEQUENCE 291 AA; 31651 MW; A241C6B9AF6D2424 CRC64;
 Query Match 17.1%; Score 88; DB 1; Length 291;
 Best Local Similarity 27.6%; Pred. No. 7.2;
 Matches 24; Conservative 21; Mismatches 24; Indels 18; Gaps 3;

QY 14 KQAEADQIIIEYIKQOVSILKEKALIQATLREKKLRVENAKLKEIEELKOE--LIQAEI 71
 Db 221 KGIILSKACDYIRE--LQTNQRMQETFEKAEIRLQMDNELLRQOIEELKNNALLRAQL 277
 QY 72 QNGVQIARPPSGTPIHANSMSENVIQ 98
 Db 278 QQ-----HNLEMGVGSSTRQ 291

Search completed: January 2, 2004, 18:21:03
 Job time : 8.47059 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:17:42 ; Search time 27.1059 Seconds
(without alignments)
1028.178 Million cell updates/sec

Title: US-09-930-169-2
Perfect score: 516
Sequence: 1 MANNDAVKRLKLEQKGAADQ.....NSMWSENVISTAVTVSSG 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virins:*
16: sp_bacteriophage:*
17: sp_archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	392	76.0	319	11 Q8C2U7	Q8C2U7 mus musculus
2	106	20.5	204	5 Q8BYX9	Q8BYX9 drosophila
3	106	20.5	323	5 Q9V504	Q9V504 drosophila
4	101	19.6	270	17 Q58907	Q58907 pyrococcus
5	96	18.6	163	17 Q58287	Q58287 pyrococcus
6	96	18.6	1900	5 Q8IMX3	Q8IMX3 drosophila
7	96	18.6	2028	5 Q9VCD1	Q9VCD1 drosophila
8	95	18.4	570	2 Q9XDC5	Q9XDC5 streptococcus
9	95	18.4	570	2 Q8NZ44	Q8NZ44 streptococcus
10	94.5	18.3	837	11 Q9JH24	Q9JH24 rat mus musculus
11	94	18.2	213	16 Q9K9R4	Q9K9R4 bacillus
12	94	18.2	609	17 Q8TXA4	Q8TXA4 methanopyru
13	93	18.0	534	2 Q38898	Q38898 streptococcus
14	93	18.0	534	2 Q68165	Q68165 streptococcus
15	92.5	17.9	1024	4 Q14527	Q14527 homo sapien
16	92.5	17.9	1046	4 Q9UDT6	Q9UDT6 homo sapien

17	92	17.8	2385	5 Q96216	Q96216 plasmidium
18	91.5	17.7	576	2 Q9KJ33	Q9KJ33 enterococcus
19	91.5	17.7	806	11 Q8VD04	Q8VD04 mus musculus
20	91	17.6	158	17 Q8U3M5	Q8U3M5 pyrococcus
21	91	17.6	2760	5 Q815T2	Q815T2 plasmidium
22	90	17.4	987	11 Q8CHF1	Q8CHF1 mus musculus
23	89.5	17.3	189	11 Q9UN83	Q9UN83 rat mus musculus
24	89	17.2	160	17 Q9UYL5	Q9UYL5 pyrococcus
25	89	17.2	816	16 Q8R7S1	Q8R7S1 thermococcus
26	89	17.2	1177	17 Q9V1R8	Q9V1R8 pyrococcus
27	88.5	17.2	673	3 Q74228	Q74228 emericella
28	88.5	17.2	718	13 Q9W6G9	Q9W6G9 xenopus lae
29	88	17.1	148	11 Q8VE59	Q8VE59 mus musculus
30	88	17.1	291	17 Q8TVS0	Q8TVS0 methanopyru
31	88	17.1	319	11 Q8EXX3	Q8EXX3 rat mus musculus
32	87.5	17.0	315	11 Q9D0Q4	Q9D0Q4 mus musculus
33	87.5	17.0	345	4 Q8TAK7	Q8TAK7 homo sapien
34	87.5	17.0	345	4 Q9UH50	Q9UH50 homo sapien
35	87.5	17.0	345	4 Q9GZM8	Q9GZM8 homo sapien
36	87.5	17.0	345	6 Q46480	Q46480 oryctolagus
37	87.5	17.0	345	11 Q9ERR1	Q9ERR1 mus musculus
38	87.5	17.0	345	11 Q9ERT6	Q9ERT6 mus musculus
39	87.5	17.0	1631	11 Q8CHG3	Q8CHG3 mus musculus
40	87	16.9	639	5 Q8MRC0	Q8MRC0 drosophila
41	87	16.9	1395	5 Q9W1R4	Q9W1R4 drosophila
42	87	16.9	2910	5 Q8IBY8	Q8IBY8 plasmidium
43	86.5	16.8	192	16 Q8R9Z3	Q8R9Z3 thermococcus
44	86.5	16.8	548	2 Q9RA74	Q9RA74 streptococcus
45	86.5	16.8	574	4 Q9H810	Q9H810 homo sapien

ALIGNMENTS

RESULT 1

Q8C2U7 Q8C2U7 PRELIMINARY; PRT; 319 AA.
ID Q8C2U7
AC Q8C2U7;
DT 01-MAR-2003 (TREMUREL_23, Created)
DT 01-MAR-2003 (TREMUREL_23, Last sequence update)
DT 01-MAR-2003 (TREMUREL_23, Last annotation update)
DE Endothelial monocyte activating polypeptide 2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
NX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NOD; TISSUE=Thymus;
RX MEDLINE=2234683; PubMed=12466851;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL: AK087932; BAC40045.1;
SQ SEQUENCE 319 AA; 35197 MW; B9778FE742BEE6 CRC64;

Qy	Query Match	76.0%; Score 392; DB 11; Length 319;
Qy	Best local similarity	76.6%; Pred. No. 7.9e-22;
Db	Matches	82; Conservative 9; Mismatches 16; Indels 0; Gaps 0;
Qy	1 MANNDAVKRLKLEQKGAADQIIETLKQVSLKREKALLOTLREKLRVBNATLKEIE 60	
Db	10 MANNDAVKRLKLEQKGAADQIIETLKQVSLKREKALLOTLREKLRVBNATLKEIE 69	
Qy	61 ELKQELIQAEIQNGVQKQIAFPSTPLHANSWSENVISTAVTVSS 107	
Db	70 ELKQELIQAEIQNGVQKQIAFPSTPLHANSWSENVISTAVTVSS 116	
RESULT 2		
Q8BYX9		

```

ID 08SYK9 PRELIMINARY; PRT; 204 AA.
AC 08SYK9;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE RE55091P.
GN CG8235.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Drenek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phoumenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY071480; AAL49102.1; -
DR FlyBase; FBgn0033351; CG8235.
SQ SEQUENCE 204 AA; 22902 MW; 3840E6034E8535B6 CRC64;

Query Match 20.5%; Score 106; DB 5; Length 204;
Best Local Similarity 31.6%; Pred. No. 1.1; Mismatches 27; Indels 4; Gaps 1;
Matches 24; Conservative 21;

QY 6 AVKRLKQKGAADQIEYLVKQVSLKKEKALIQATLREKKLRVENAKLKKEIEELKOE 65
DB 31 ADLQQLASNNRERAEALINSIEAISIGIQ-----QLVERKQGEILIKENALAKEVEALAQ 86
QY 66 LIAQETIONGVKQIAPF 81
DB 87 LVQLEIRNGKQIAPF 102
DB 87 LVQLEIRNGKQIAPF 102

RESULT 3
QY 09VS04 PRELIMINARY; PRT; 323 AA.
AC 09VS04;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE CG8235 protein.
GN CG8235.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA MEDLINE=201936006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,
RA Baillew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotcher P.,
RA Burks K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke S., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson R., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

```

```

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milehina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN 12
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banon J., An H., Baldwin D., Banon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson R., Dorsett V., Doup L.E., Doyle C., Drenek D., Farfan D.,
RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phoumenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirek R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A.
RA Mista S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminiker J.S., Prochuk S.E., Smith C.D.,
RA Tuhy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN 14
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN 15
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AEO01835; AAF59019.2; -
DR FlyBase; FBgn0033351; CG8235.
DR InterPro; IPR002547; tRNA bind.
DR Pfam; PF01588; tRNA bind.1.
SQ SEQUENCE 323 AA; 34401 MW; 6EBB6E30115FAE84 CRC64;

Query Match 20.5%; Score 106; DB 5; Length 323;
Best Local Similarity 31.6%; Pred. No. 1.8; Mismatches 27; Indels 4; Gaps 1;
Matches 24; Conservative 21;

QY 6 AVKRLKQKGAADQIEYLVKQVSLKKEKALIQATLREKKLRVENAKLKKEIEELKOE 65
DB 31 ADLQQLASNNRERAEALINSIEAISIGIQ-----QLVERKQGEILIKENALAKEVEALAQ 86
QY 66 LIAQETIONGVKQIAPF 81

```

Db 87 LVQLELRNGKQIPIVP 102

RESULT 4

058907 ID 058907 PRELIMINARY; PRT; 270 AA.

AC 058907; STRAIN=OT3;

DT 01-AUG-1998 (TREMBlrel. 07, Created)

DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)

DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE Hypothetical protein PH189.

GN PH189.

OS Pyrococcus horikoshii.

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;

OC Pyrococcus.

OX NCBI_TaxID=53953;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OT3;

RX MEDLINE=98344137; PubMed=9679194;

RA Kawarabayashi Y., Sawada M., Horikawa H., Hainaka Y., Hino Y., Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohtuku Y., Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K., Mauchly Y., Shizuya H., Kikuchi H.;

RA "Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";

RT DNA Res. 5:55-76(1998).

RL EMBL; AP000005; BAA30289.1;

KW Hypothetical protein; Complete proteome.

SO SEQUENCE 270 AA; 31019 MW; AA5A4C8B94FB1EFE CRC64;

Query Match 19.6%; Score 101; DB 17; Length 270;

Best Local Similarity 30.0%; Pred. No. 3.5;

Matches 30; Conservative 22; Mismatches 32; Indels 16; Gaps 3;

058287 ID 058287 PRELIMINARY; PRT; 163 AA.

AC 058287; STRAIN=OT3;

DT 01-AUG-1998 (TREMBlrel. 07, Created)

DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Hypothetical protein PH0552.

GN PH0552.

OS Pyrococcus horikoshii.

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;

OC Pyrococcus.

OX NCBI_TaxID=53953;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OT3;

RX MEDLINE=98344137; PubMed=9679194;

RA Kawarabayashi Y., Sawada M., Horikawa H., Hainaka Y., Hino Y., Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohtuku Y., Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K., Mauchly Y., Shizuya H., Kikuchi H.;

RA "Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";

RT DNA Res. 5:55-76(1998).

RL EMBL; AP000005; BAA30289.1;

KW Hypothetical protein; Complete proteome.

SO SEQUENCE 270 AA; 31019 MW; AA5A4C8B94FB1EFE CRC64;

DR EMBL; AP000002; BAA29641.1;

DR InterPro; IPR004883; DUF260.

DR Pfam; PF03195; DUF260; 1.

KW Hypothetical protein; Complete proteome.

SO SEQUENCE 163 AA; 19266 MW; 86C178CF9CF2EBSF CRC64;

Query Match 18.6%; Score 96; DB 17; Length 163;

Best Local Similarity 33.3%; Pred. No. 5;

Matches 24; Conservative 19; Mismatches 21; Indels 8; Gaps 2;

058907 ID 058907 PRELIMINARY; PRT; 1900 AA.

AC 058907; STRAIN=OT3;

DT 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE CG6129-PA.

GN CG6129-PA.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=20196006; PubMed=10731132;

RX Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L., Abail J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktarglu L., Beasley E.M., Beeson K.Y., Berens P.V., Berman B.P., Bhandari D., Bolintsov S., Borokova D., Botchan M.R., Bouck J., Brocks P., Brotter P., Burlingame K.C., Buzam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Fowler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibeagwa C., Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclab J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H., Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier B.E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J., Williams S.M., Woodruff W., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

Science 287:2185-2195(2000).

[2]

SEQUENCE FROM N.A.

RA Ceiniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,

RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busan D.A.,

RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,

RA Dodson K., Dorett V., Doup L.B., Doyle C., Dresnek D., Farfan D.,

RA Fernandez S., Frise E., Galle R.F., Garg N.S., George R.A.,

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,

RA Idegaw C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,

RA Pacleb U., Paragas V., Park S., Patel S., Pfeiffer B.,

RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,

RA Stapleton M., Strong R., Svirekas R., Tector C., Tyler D.,

RA Williams S.M., Zaveri J.S., Smith H.O., Venter C.J., Rubin G.M.,

RT "Sequencing of Drosophila melanogaster genome."

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,

RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Ceiniker S.E.,

RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,

RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,

RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,

RA Ashburner M., Gelbart W.M., Rubin G.M., Mangali C.J., Lewis S.E.,

RT "Annotation of Drosophila melanogaster genome."

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RA Adams M.D., Ceiniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RA FlyBase;

RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: AEO03746; AAI13982.1; -

SO SEQUENCE 1900 AA; 215425 MW; 3BD3CE23B03002A CAC64;

Query Match 18.6%; Score 96; DB 5; Length 1900;

Best Local Similarity 26.1%; Pred. No. 54;

Matches 36; Conservative 15; Mismatches 41; Indels 46; Gaps 4;

QY 11 LKQKGAADPITRYKQVVS-----LKKKX-----I 37

DB 679 LEOQRIESDNLINLEKQKSDLEYDLDKLLKCDLQEKHEKLSNNSCSTDELKSVQNC 738

QY 38 LQATLEBEKRLRYENAKLKEIEELKQEL-----IOAEIONGVQVQIAFPSCGPLHANS 90

DB 739 LQGAQERKRLRQSDVQNCIEIGELKKEALIDKARLELETDN-----LSAEXKLCIQ 792

QY 91 MASENVIQSTAVTTSVG 108

DB 793 LKEXKILQDLACVTRDRG 810

RESULT 7

QVCDI PRELIMINARY; PRT; 2028 AA.

AC QVCDI;

DT 01-MAY-2000 (T-EMBLrel. 13, Created)

DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)

DE CG6129 protein.

GN CG6129.

OS Drosophila melanogaster (fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephyroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkelley;

RA MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,

RA Abil J.F., Achayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhanderi D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottler P.,

RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Dugan-Rocha S., Dunker B.C., Dunn P.,

RA Dodson K., Doup L.B., Domes M., Dugan-Rocha S., Fleschmann W.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaser K.,

RA Glodex A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegaw C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulj D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Mlshina N.V., Mobarri C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Murzy D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb U.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,

RA Svirekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach U.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter C.J.,

RT "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195(2000).

RN

RP SEQUENCE FROM N.A.

RA Ceiniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,

RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busan D.A.,

RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,

RA Dodson K., Dorett V., Doup L.B., Doyle C., Dresnek D., Farfan D.,

RA Fernandez S., Frise E., Galle R.F., Garg N.S., George R.A.,

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,

RA Idegaw C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,

RA Pacleb U., Paragas V., Park S., Patel S., Pfeiffer B.,

RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,

RA Stapleton M., Strong R., Svirekas R., Tector C., Tyler D.,

RA Williams S.M., Zaveri J.S., Smith H.O., Venter C.J., Rubin G.M.,

RT "Sequencing of Drosophila melanogaster genome."

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,

RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Ceiniker S.E.,

RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,

RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,

RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,

RA Ashburner M., Gelbart W.M., Rubin G.M., Mangali C.J., Lewis S.E.,

RT "Annotation of Drosophila melanogaster genome."

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RA Adams M.D., Ceiniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.
RA FlyBase:
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases
DR EMBL: AE003746; AAF56238.2; -
DR FlyBase: FBgn0039152; CG6129.
SQ SEQUENCE 2028 AA; 230473 MW; 7CE8C30565A31A57 CFC64

Query Match	18.6%;	Score 96;	DB 5;	Length 2028;
Best Local Similarity	26.1%;	Pred. No. 58;		
Matches	36;	Conservative	15;	Mismatches 41;
				Indels 46;
				Gaps 4

```

QY      11  LEKGGAADIIIEYUQAQS-----LYKKA-----I 37
      ||| : : : : : |||
Db      679  LEQGRFESDMLINLFRKQSDLEUYDIDKILYKCOLQEKHEKLSNNSGTSDELKSYQNC 738
      ||| : : : : : |||

QY      38  LQATLEEKYLRPYNAKKKEIYEIKQEL-----IQAEIONGYKQAFPSGPTPLANS 90
      ||| : : : : : |||
Db      739  LQAEQERKRYLRQSDYQDCNEIGELKQELAILDKARLEETDN-----LSAGEKICLQ 792
      ||| : : : : : |||

QY      91  MASENNVIOSTAVTIVSSG 108
      : : : : :
Db      793  LEKEKILQDLACVTRDNG 810
      : : : : :

```

RESULT	ID	Q9XDC5	PRELIMINARY;	PRT;	570 AA.
09XDC5	09XDC5				
DT	01-NOV-1999	(TrEMBLrel. 12, Created)			
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)			
DT	01-MAR-2003	(TrEMBLrel. 23, Last annotation update)			
DE	Protective antigen.				
GN	SPAL8.				
OS	Streptococcus pyogenes.				
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;				
OC	Streptococcus.				
OX	NCBI_TaxID=1314;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RP	STRAIN=87-282;				
RX	MEDLINE=99242623; PubMed=10225969;				
RA	Dale J.B., Chiang E.Y., Liu S., Courtney H.S., Hasty D.L.;				
RT	"New protective antigen of group A streptococci.";				
RL	J. Clin. Invest. 103:1261-1268 (1999).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RP	STRAIN=87-282;				
RX	MEDLINE=21189206; PubMed=11292710;				
RA	McLellan D.S.U., Chiang E.Y., Courtney H.S., Hasty D.L., Wei S.C.,				
RT	Hu M.C., Walls M.A., Bloom J.U., Dale U.B.;				
RL	"Spa contributes to the virulence of type 18 group A streptococci.";				
CC	Infect. Immun. 69:2943-2949 (2001).				
CC	-1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY				
CC	AN AMIDE BOND (BY SIMILARITY).				
EMBL	AF086813; AAD42939.2; -				
DR	InterPro: IPR005877; Gpos_Y08R.				
DR	InterPro: IPR001899; Gram_pos_anchor.				
DR	InterPro: IPR006192; LPXTG.				
DR	Pfam: PF00746; Gram_pos_anchor. 1.				
DR	Pfam: PF04650; YSIRK signal. 1.				
DR	PRINTS: PR00015; GPOSANCHOR.				
DR	TIGRFAMs: TIGR01167; LPXTG_anchor. 1.				
DR	TIGRFAMs: TIGR01168; YSIRK_signal. 1.				
DR	PROSITE: PS50847; GRAM_POS_ANCHORING. 1.				
DR	Cell wall; peptidoglycan-anchor.				
KQ	SEQUENCE 570 AA; 62718 MW; 15D27159422D1E50 CRC64;				

```
Query Match      18.4%; Score 95; DB 2; Length 570;
Match Similarity 29.2%; Pred. No. 20;
Matches          40; Conservative 16; Indels 30; Gaps 4.
```

```

Db 421 AASDAKVAIEKEVEAAKAEVADLKAQAKKEBEELEAVKEKEALKEIKELKKAHAEEL 480
Qy 52 -----NAKKEIEELKEOEI-----IOAEIONGVQKQIAFPGSTPYLHNSM----- 91
Db 481 SKUKLEKEDKHANADLOAEIKRKLKQELADRIKLSQSGRASQYNPOSTTAKKAGQLPSTG 540
Qy 92 VSENVIOSTAVTVTSSG 108
Db 541 ESANPEFTTAAITVING 557

```

RESULT	9
Q8NZ44	Q8NZ44
ID	Q8NZ44
AC	Q8NZ44
DT	01-OCT-2002 (TREMBLrel. 22, Created)
DT	01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE	Streptococcal protective antigen.
GN	SPA OR SPYM18_2046.
OS	Streptococcus pyogenes (serotype M18).
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC	Streptococcus.
OX	NCBI_TaxID=186103;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=MGAS8232 / Serotype M18;
RX	MEDLINE=21927593; PubMed=11917108;
RA	Smoot J.C., Barbhan K.D., Van Gompel J.T., Smoot L.M., Chaussee M.S.,
RA	Sylvia G.L., Sturdevant D.E., Ricklets S.M., Forcella S.F.,
RA	Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA	Kapur V., Daly J.A., Veasy L.G., Musser J.M.,
RT	"Genome sequence and comparative microarray analysis of serotype M18
RT	group A Streptococcus strains associated with acute rheumatic fever
RT	outbreaks".
RL	Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
DR	EMBL; AE010108; AAL98520.1; -.
DR	InterPro; IPR005877; Gpos_YISRK.
DR	InterPro; IPR001899; Gpos_pos_anchor.
DR	InterPro; IPR006192; LpYTG.
DR	Pfam; PF00746; Gram_pos_anchor; 1.
DR	Pfam; PF04650; YISRK_signal; 1.
DR	PRINTS; PR00015; GPOSANCHOR.
DR	TIGRFAMs; TIGR01167; LpYTG_anchor; 1.
DR	TIGRFAMs; TIGR01168; YISRK_signal; 1.
DR	PROSITE; P650847; GRAM_POS_ANCHORING; 1.
DR	Complete proteome.
Q8	SEQUENCE 570 AA; 62737 MW; 8CD65B66B1E1D6C CRC64;

Query Match 18.4%; Score 95; DB 16; Length 570;
Best Local Similarity 29.2%; Pred. No. 20;
Matches 40; Conservative 16; Mismatches 51; Indels 30; Gaps 4;

[illegible]

RESULT	ID	PRELIMINARY	PRT
09JH24	09JH24		837 AA.
AC	09JH24		
DT	01-OCT-2000	(17EMBRel. 15.	Created)
DT	01-OCT-2000	(17EMBRel. 15.	Last sequence update)
DT	01-OCT-2002	(17EMBRel. 22.	Last annotation update)

DE GRIP-associated protein 1 long form.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley;
 RX MEDLINE=20353053; PubMed=10896157;
 RA Ye B., Liao D., Zhang X., Zhang P., Dong H., Hugenir R.L.;
 RT "GRASP-1: A Neuronal RasGEF Associated with the AMPA Receptor/GRIP
 Complex.";
 RL Neuron 26:603-617(2000).
 DR EMBL: AF274057; AAF82298.1; -.
 SQ SEQUENCE 837 AA; 96074 MW; A746AEAFD09D3AD2 CRC64;

Query Match 18.2%; Score 94.5; DB 11; Length 837;
 Best Local Similarity 27.8%; Pred. No. 32;
 Matches 32; Conservative 20; Mismatches 48; Indels 15; Gaps 3;

QY 1 MANNDAVLKRLBOKGADQIIEYLKQVS-----LLKKAIIQATLR-BEKK 47
 DB 31 LRKNGVELSLRQVAVLDR--EFSKQKALSKSKAQEVEVLSEKEMIQAKLHSGEED 88
 QY 48 LRVENAKLKEIEBELKQELIOAEIQNGVKQIAPPSGTPHLANSWSENVIQSTAV 102
 DB 89 FRIQNSTLMAFSLCSQLQLELNNQLKEGVPGAGPRVDELLRLQMENTAL 143

RESULT 11
 ID Q9K9R4 PRELIMINARY; PRT; 213 AA.

AC Q9K9R4;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Hypothetical protein BH2581.
 GN BH2581.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Maui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 DE "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL: AP001516; BAB06300.1; -.
 DR InterPro: IPR001005; MyD_DNA_binding.
 DR SMART: SM00717; SANT; 1.
 DR PROSITE: PS00037; MYB_1; 1.
 DR PROSITE: PS50090; MYB_3; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 213 AA; 24524 MW; D2C04E7F43673BB6 CRC64;

Query Match 18.2%; Score 94; DB 16; Length 213;
 Best Local Similarity 36.8%; Pred. No. 9.2;
 Matches 25; Conservative 16; Mismatches 15; Indels 12; Gaps 3;

QY 7 VLKLEBQKAE--ADQIIEYLKQVSIL-----KKKAIQATLRBEKKLRVENAKLKK 57
 DB 115 VSSGVNDNGSGITWEKVIAPLKNQOTLLGODRLQKCK--QQLDENETLRKNNALEK 171
 QY 58 EIEELKOE 65
 DB 172 EIKKROE 179

RESULT 12

Q8TXA4
 ID Q8TXA4 PRELIMINARY; PRT; 609 AA.
 AC Q8TXA4;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Uncharacterized protein.
 GN MK0771.
 OS Methanopyrus kandleri.
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyracaceae;
 OC Methanopyrus.
 NCBI_TaxID=2320;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AV19 / DSM 6324 / JCM 9639;
 RX MEDLINE=21927647; PubMed=11930014;
 RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
 RA Shcherbina O.V., Shakhova V.V., Belova G.I., Aravind L.,
 RA Natale D.A., Rogozin I.B., Tatunov R.L., Wolf Y.I., Stetter K.O.,
 RA Malykh A.G., Koonin E.V., Kozlovskii S.A.;
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
 and monophyly of archaeal methanogens.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
 DR EMBL: AB010369; AA01985.1; -.
 DR HSSP: P04268; IIC2.
 DR InterPro: IPR002017; Spectrin.
 KW Complete proteome.
 SQ SEQUENCE 609 AA; 69552 MW; AB10C9780DC5AD78 CRC64;

Query Match 18.2%; Score 94; DB 17; Length 609;
 Best Local Similarity 32.6%; Pred. No. 25;
 Matches 31; Conservative 21; Mismatches 23; Indels 20; Gaps 4;

QY 1 MANNDAVLKRLBOK-----GAEDQIIEYLKQVSLLKKAIIQATLRBEKK----- 47
 DB 209 LAEN--LKKLKEKNEIEERDRLEETKEVKGKLQKQAKLQSKLKEVSRDLANV 265
 QY 48 --LRVENAKLKEIEBELKQEL--IOAEIQNGVKQI 78
 DB 266 EALRNENETLRKKIDKLSLNLQKLDREKRL 300

RESULT 13

ID Q33898 PRELIMINARY; PRT; 534 AA.

AC Q33898;
 DT 01-JUN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE M-protein.
 GN SEM.
 OS Streptococcus equi.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 NCBI_TaxID=1336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Timoney J.F., Artushin S.C.;
 RT "The M-protein of Streptococcus equi.";
 RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOLYCAN BY
 AN AMIDE BOND (BY SIMILARITY).
 DR EMBL: U73162; AAB71984.1; -.
 DR InterPro: IPR005877; Gpos_YsIRK.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR InterPro: IPR006192; LPXTG.
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR Pfam: PF04650; YsIRK_signal; 1.
 DR PRINTS: PR00015; GPOSANCHOR.
 DR TIGRPFAM: TIGR01167; LPXTG_anchor; 1.
 DR TIGRPFAM: TIGR01168; YsIRK_signal; 1.
 DR PROSITE: PS50847; GRAM_POS_ANCHORING; 1.
 KW Cell wall; Peptidoglycan-anchor.

SQ SEQUENCE 534 AA; 58213 MW; 654A4AB90E282801 CRC64;

Query Match 18.0%; Score 93; DB 2; Length 534;
Best Local Similarity 29.2%; Pred. No. 27;
Matches 40; Conservative 16; Mismatches 51; Indels 30; Gaps 4;

QY 2 ANNDVLRLEKLEKGAADQIIEYKQV-----SILKEKAILQATLREKKLRVE-- 51
DB 385 AASDAKVALEKEVEAAKAVADLKQAQAKKEELEAVKEKEALEAKIEELKKAHAHEL 444

QY 52 -----NAKKEIEELKQEL---IQAEIQNGVQIAPSGTPIHANSM---- 91
DB 445 SKLKEMLEKKDHANADIQAEINRLKQELADRIKSLSGGRASQTNPGTTAKAQLPSTG 504

QY 92 VSENVIOSTAVTVSSG 108
DB 505 ESANPFTIALTVIAG 521

RESULT 14

ID 068165 PRELIMINARY; PRT; 534 AA.

AC 068165.
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Fibrinogen-binding protein.
GN FIB.

OS Streptococcus equi.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1336;
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=TW.

RA MEDLINE=96240236; PubMed=9579073;

RX Meehan M., Nowlan P., Owen P.;

RT "Affinity purification and characterization of a fibrinogen-binding

RT protein complex which protects mice against lethal challenge with

RT Streptococcus equi subsp. equi";

RL Microbiology 144;993-1003(1998).

CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY

CC AN AMIDE BOND (BY SIMILARITY).

CC EMBL; AF012927; AAC38445.1; -.

DR InterPro; IPR005877; Gpos_YsIRK.

DR InterPro; IPR001899; Gram_pos_anchor.

DR InterPro; IPR006192; LPXTG.

DR Pfam; PF00746; Gram_pos_anchor; 1.

DR Pfam; PF04650; YsIRK signal; 1.

DR PRINTS; PR00015; GPOSANCHOR.

DR TIGRPFAMs; TIGR01167; LPXTG_anchor; 1.

DR TIGRPFAMs; TIGR01168; YsIRK_signal; 1.

DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.

KW Cell wall; Peptidoglycan-anchor.

SQ SEQUENCE 534 AA; 58344 MW; A8995371273D2B1A CRC64;

Query Match 18.0%; Score 93; DB 2; Length 534;
Best Local Similarity 29.9%; Pred. No. 27;

Matches 41; Conservative 15; Mismatches 51; Indels 30; Gaps 5;

QY 2 ANNDVLRLEKLEKGAADQIIEYKQVSL---LKEKAILQATLREKKLRVE-- 51
DB 385 AASDAKVALEKEVEAAKAVADLKQAQAKKEELEAVKEKEALEAKIEELKKAHAHEL 444

QY 52 -----NAKKEIEELKQEL---IQAEIQNGVQIAPSGTPIHANSM---- 91
DB 445 SKLKEMLEKKDHANADIQAEINRLKQELADRIKSLSGGRASQTNPGTTAKAQLPSTG 504

QY 92 VSENVIOSTAVTVSSG 108
DB 505 ESANPFTIALTVIAG 521

RESULT 15

ID 014527 PRELIMINARY; PRT; 1024 AA.

AC 014527.
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical protein KIAA0291 (Fragment).
GN KIAA0291.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RA Ohara O., Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N.,

RA Nomura N.;

RT "Prediction of the coding sequences of unidentified human genes.";

RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB006629; BAA22960.2; -.

DR InterPro; IPR000938; CAP-Gly.

DR Pfam; PF01302; CAP_GLY; 2.

DR PROSITE; PS00845; CAP_GLY_1; 2.

DR PROSITE; PS50245; CAP_GLY_2; 2.

KW Hypothetical protein.

FT NON TER 1

SQ SEQUENCE 1024 AA; 112981 MW; 12278FC2678B8620 CRC64;

Query Match 17.9%; Score 92.5; DB 4; Length 1024;
Best Local Similarity 26.9%; Pred. No. 54;

Matches 25; Conservative 30; Mismatches 31; Indels 7; Gaps 2;

QY 8 LKRLKQGAADQIIEYKQVSL---LKEKAILQATLREKKLRVENAKKEIEELKQELI 67
DB 713 LKRLKQGAADQIIEYKQVSL---LKEKAILQATLREKKLRVENAKKEIEELKQELI 766

QY 68 QAEIO-NGVQIAPSGTPIHANSVSENVIOS 99
DB 767 VAEINRLQAEIALCSQHTMIESNDISEETIRT 799

Search completed: January 2, 2004, 18:22:21
Job time : 28.1059 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:19:07 ; Search time 12.2824 Seconds
(without alignments)
372.044 Million cell updates/sec

Title: US-09-930-169-2
Perfect score: 516
Sequence: 1 MANDAVLKRLKQKGAADQ.....NSMSENVIOSTAVTVSSG 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/PTCUTS.COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	511	99.0	312	US-08-360-821B-36	Sequence 36, Appl
2	500.5	97.0	310	US-08-129-456A-37	Sequence 37, Appl
3	395	76.6	310	US-08-129-456A-36	Sequence 36, Appl
4	395	76.6	310	US-08-705-868-3	Sequence 3, Appl
5	395	76.6	310	US-09-123-615-3	Sequence 3, Appl
6	395	76.6	310	US-08-360-821B-35	Sequence 35, Appl
7	93	18.0	534	US-09-103-664A-2	Sequence 2, Appl
8	90	17.4	956	US-09-914-259-17	Sequence 17, Appl
9	90	17.4	957	US-09-914-259-16	Sequence 16, Appl
10	89.5	17.3	815	US-09-914-259-18	Sequence 18, Appl
11	88	17.1	667	US-09-071-709-9	Sequence 9, Appl
12	87.5	17.0	459	US-09-071-709-1	Sequence 1, Appl
13	84	16.3	741	US-09-854-856-60	Sequence 60, Appl
14	84	16.3	769	US-09-854-856-44	Sequence 44, Appl
15	84	16.3	801	US-09-854-856-28	Sequence 28, Appl
16	84	16.3	829	US-09-854-856-12	Sequence 12, Appl
17	84	16.3	829	US-09-854-856-54	Sequence 54, Appl
18	84	16.3	922	US-09-854-856-38	Sequence 38, Appl
19	84	16.3	954	US-09-854-856-22	Sequence 22, Appl
20	84	16.3	982	US-09-854-856-6	Sequence 6, Appl
21	84	16.3	1911	US-09-854-856-64	Sequence 64, Appl
22	84	16.3	1939	US-09-854-856-48	Sequence 48, Appl
23	84	16.3	1971	US-09-854-856-32	Sequence 32, Appl
24	84	16.3	1999	US-09-854-856-16	Sequence 16, Appl
25	84	16.3	2004	US-09-854-856-58	Sequence 58, Appl
26	84	16.3	2032	US-09-854-856-42	Sequence 42, Appl
27	84	16.3	2048	US-09-854-856-62	Sequence 62, Appl

28	84	16.3	2064	US-09-854-856-26	Sequence 26, Appl
29	84	16.3	2076	US-09-854-856-46	Sequence 46, Appl
30	84	16.3	2092	US-09-854-856-10	Sequence 10, Appl
31	84	16.3	2108	US-09-854-856-30	Sequence 30, Appl
32	84	16.3	2136	US-09-854-856-14	Sequence 14, Appl
33	84	16.3	2141	US-09-854-856-56	Sequence 56, Appl
34	84	16.3	2157	US-09-854-856-52	Sequence 52, Appl
35	84	16.3	2169	US-09-854-856-40	Sequence 40, Appl
36	84	16.3	2185	US-09-854-856-36	Sequence 36, Appl
37	84	16.3	2201	US-09-854-856-24	Sequence 24, Appl
38	84	16.3	2217	US-09-854-856-20	Sequence 20, Appl
39	84	16.3	2229	US-09-854-856-8	Sequence 8, Appl
40	84	16.3	2245	US-09-854-856-4	Sequence 4, Appl
41	84	16.3	2294	US-09-854-856-50	Sequence 50, Appl
42	84	16.3	2322	US-09-854-856-34	Sequence 34, Appl
43	84	16.3	2354	US-09-854-856-18	Sequence 18, Appl
44	84	16.3	2382	US-09-854-856-2	Sequence 2, Appl
45	83	16.1	414	PCT-US93-03077-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-360-821B-36
Sequence 36, Application US/08360821B
Patent No. 6228837
GENERAL INFORMATION:
APPLICANT: Stern, David M.
APPLICANT: Clauss, Matthias
APPLICANT: Kao, Janet
APPLICANT: Kayton, Mark
APPLICANT: Libutti, Steven K
TITLE OF INVENTION: Endothelial Monocyte Activating Polypeptide II: A Mediator Which Activates Host Response
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham, LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.30, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/360, 821B
FILING DATE: 08-OCT-96
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28, 678
REFERENCE/DOCKET NUMBER: 41735-A-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 312 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-360-821B-36
Query Match 99.0%; Score 511; DB 3; Length 312;
Best Local Similarity 99.1%; Pred. No. 2.4e-45;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MANDAVLKRLKQKGAADQIIIEYLKQGVSLKKEKALIQATLREBKRLRVENAKLKEIE 60

Db 1 MANDAVLRLKLEKGAADQIIEYLKQVSLKKEKALQATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIQAIEIONGVQVQIAFPSSGTPHLANSMSENVIOSTAVTTVSSG 108
Db 61 ELKQELIQAIEIONGVQVQIAFPSSGTPHLANSMSENVIOSTAVTTVSSG 108

RESULT 2

US-08-129-456A-37
Sequence 37, Application US/08129456A
Patent No. 5641867
GENERAL INFORMATION:
APPLICANT: Stern, David M.
APPLICANT: Claus, Matthias
APPLICANT: Kao, Janet
APPLICANT: Kayton, Mark
APPLICANT: Libutti, Steven K.
TITLE OF INVENTION: Endothelial-Monocyte Activating
TITLE OF INVENTION: Polypeptide II: A Mediator Which
TITLE OF INVENTION: Activates Host Response
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0 Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/129,456A
FILING DATE: 29-SEP-1993
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41735
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 278 0400
TELEFAX: 212 391 0526
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-129-456A-37

Query Match 97.0%; Score 500.5; DB 1; Length 310;
Best Local Similarity 98.2%; Pred. No. 2.9e-44;
Matches 107; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 MANDAVLRLKLEKGAADQIIEYLKQVSLKKEKALQATLREKKLRVENAKLKEIE 60
Db 1 MANDAVLRLKLEKGAADQIIEYLKQVSLKKEKALQATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIQAIEIONGVQVQIAFPSSGTPHLANSMSENVIOSTAVTTVSSG 108
Db 61 ELKQELIQAIEIONGVQVQIAFPSSGTPHLANSMSENVIOSTAVTTVSSG 108

RESULT 3

US-08-129-456A-36
Sequence 36, Application US/08129456A
Patent No. 5641867
GENERAL INFORMATION:
APPLICANT: Stern, David M.
APPLICANT: Claus, Matthias

APPLICANT: Kao, Janet
APPLICANT: Kayton, Mark
APPLICANT: Libutti, Steven K.
TITLE OF INVENTION: Endothelial-Monocyte Activating
TITLE OF INVENTION: Polypeptide II: A Mediator Which
TITLE OF INVENTION: Activates Host Response
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0 Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/129,456A
FILING DATE: 29-SEP-1993
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41735
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 278 0400
TELEFAX: 212 391 0526
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-129-456A-36

Query Match 76.6%; Score 395; DB 1; Length 310;
Best Local Similarity 77.6%; Pred. No. 2.6e-33;
Matches 83; Conservative 8; Mismatches 16; Indels 0; Gaps 0;
QY 1 MANDAVLRLKLEKGAADQIIEYLKQVSLKKEKALQATLREKKLRVENAKLKEIE 60
Db 1 MANDAVLRLKLEKGAADQIIEYLKQVSLKKEKALQATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIQAIEIONGVQVQIAFPSSGTPHLANSMSENVIOSTAVTTVSS 107
Db 61 ELKQELIQAIEIONGVQVQIAFPSSGTPHLANSMSENVIOSTAVTTVSS 107

RESULT 4

US-08-705-868-3
Sequence 3, Application US/08705868
Patent No. 5885798
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Coleman, Roger
APPLICANT: Au-Young, Janice
APPLICANT: Murty, Lynn E.
TITLE OF INVENTION: NOVEL MONOCYTE ACTIVATING CYTOKINE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/705,868
FILING DATE: Filed Herewith
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0117 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 498910
US-08-705-868-3

Query Match 76.6%; Score 395; DB 2; Length 310;
Best Local Similarity 77.6%; Pred. No. 2.6e-33;
Matches 83; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

Qy 1 MANDAVAKRLKQKGAEDQIIETLKQVSLKKEKALIQATLRREKKLRVENAKLKEIE 60
Db 1 MANDAVAKRLKQKGAEDQIIETLKQVSLKKEKALIQATLRREKKLRVENAKLKEIE 60
Qy 61 ELKQELIAEIQNGVKQIAFPSTGPLHANSWSENVIOSTAVTVSS 107
Db 61 ELKQELIAEIQNGVKQIAFPSTGPLHANSWSENVIOSTAVTVSS 107

RESULT 5
US-09-123-615-3
Sequence 3, Application US/09123615
Patent No. 6090377
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Coleman, Roger
APPLICANT: Au-Young, Janice
APPLICANT: Murry, Lynn E.
TITLE OF INVENTION: NOVEL MONOCYTE ACTIVATING CYTOKINE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/123,615
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,868
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0117 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 498910
US-09-123-615-3

Query Match 76.6%; Score 395; DB 3; Length 310;
Best Local Similarity 77.6%; Pred. No. 2.6e-33;
Matches 83; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

Qy 1 MANDAVAKRLKQKGAEDQIIETLKQVSLKKEKALIQATLRREKKLRVENAKLKEIE 60
Db 1 MANDAVAKRLKQKGAEDQIIETLKQVSLKKEKALIQATLRREKKLRVENAKLKEIE 60
Qy 61 ELKQELIAEIQNGVKQIAFPSTGPLHANSWSENVIOSTAVTVSS 107
Db 61 ELKQELIAEIQNGVKQIAFPSTGPLHANSWSENVIOSTAVTVSS 107

RESULT 6
US-08-360-821B-35
Sequence 35, Application US/08360821B
Patent No. 6228837

GENERAL INFORMATION:
APPLICANT: Stern, David M.
APPLICANT: Klaus, Matthias
APPLICANT: Kao, Janet
APPLICANT: Libutti, Steven K
TITLE OF INVENTION: Endothelial Monocyte Activating
TITLE OF INVENTION: Polypeptide II: A Mediator which Activates Host Response
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham, LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.30, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/360,821B
FILING DATE: 08-OCT-96
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41735-A-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-360-821B-35

Query Match 76.6%; Score 395; DB 3; Length 310;
Best Local Similarity 77.6%; Pred. No. 2.6e-33;
Matches 83; Conservative 8; Mismatches 16; Indels 0; Gaps 0;


```

: CITY: PALO ALTO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/071,709
: FILING DATE: Filed Herewith
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: CERRONE, MICHAEL C.
: REGISTRATION NUMBER: 39,132
: REFERENCE/DOCKET NUMBER: PF-0513 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (650) 855-0555
: TELEFAX: (650) 845-4166
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 459 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: THPINOBI
: CLONE: 031381
US-09-071-709-1

Query Match          17.0%; Score 87.5; DB 3; Length 459;
Best Local Similarity 26.8%; Pred.No. 0.33;
Matches 30; Conservative 24; Mismatches 45; Indels 13; Gaps 4;

Qy      8 LKRLPEQGAGADQ-----ITLKQVS-LKEKRIILQATIREKKLAVENAKLKET 59
       ::|||::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Db      115 VRELQANDLERAKRATIVSLBPFQRLNQALERNAFLESELDKESLLVSVORLDEA 174
        .|.||||.||..||..||..||..||..||..||..||..||..||..||..||..||
Qy      60 ELKQELIQAEIONGVQIAPSGTPIHANSWSENVI--STAATTSSG 108
       .|.||||.||..||..||..||..||..||..||..||..||..||..||..||..||
Db      175 RDLQELAVREQQEYTRKSAPSPTIDCEKM--DSAVQASLSLPATPVKG 224
        .|.||||.||..||..||..||..||..||..||..||..||..||..||..||..||

RESULT 13
US-09-854-856-60
: Sequence 60, Application US/09854856
: Patent No. 6541252
: GENERAL INFORMATION:
: APPLICANT: Walke, D. Wade
: APPLICANT: Hilbun, Erin
: APPLICANT: Donoho, Gregory
: APPLICANT: Turner, C. Alexander Jr.
: TITLE OF INVENTION: Encoding the Same
: FILE REFERENCE: LEX-0178-USA
: CURRENT APPLICATION NUMBER: US/09/854,856
: CURRENT FILING DATE: 2001-05-14
: PRIOR APPLICATION NUMBER: US 60/206,015
: PRIOR FILING DATE: 2000-05-19
: NUMBER OF SEQ ID NOS: 64
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 60
: LENGTH: 741
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (1)..(741)
: OTHER INFORMATION: Xaa = Any Amino Acid
US-09-854-856-60

```

Best Local Similarity 34.9%; Pred. No. 1.4;
Matches 30; Conservative 13; Mismatches 19; Indels 24; Gaps 5;

QY 25 LKQOVSLLKEKAILQATLREKKL-RVENAKLKEIELEKOEILQAEI-ONGVKQIAPFS 82
Db 498 IKDVSLLIKRK-----RQROLVREOEKKQESSLSKQOVQSSASQTGIKQLPSAS 550

QY 83 -GTPPLHNSMVSENVIOSTAVTTVSS 107
Db 551 TGIP-----TASTTSAS 562

RESULT 14
US-09-854-856-44
; Sequence 44, Application US/09854856
; Patent No. 6541252

; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade

; APPLICANT: Hilbun, Erin

; APPLICANT: Donoho, Gregory

; TITLE OF INVENTION: Turner, C. Alexander Jr.
TITLE OF INVENTION: Encoding the Same

FILE REFERENCE: LEX-0178-USA

CURRENT APPLICATION NUMBER: US/09/854,856

CURRENT FILING DATE: 2001-05-14

PRIOR APPLICATION NUMBER: US 60/206,015

PRIOR FILING DATE: 2000-05-19

NUMBER OF SEQ ID NOS: 64

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 44

LENGTH: 769

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: VARIANT

LOCATION: (1)...(769)

OTHER INFORMATION: Xaa = Any Amino Acid

US-09-854-856-44

Query Match

Best Local Similarity 16.3%; Score 84; DB 4; Length 769;

Matches 30; Conservative 13; Mismatches 19; Indels 24; Gaps 5;

QY 25 LKQOVSLLKEKAILQATLREKKL-RVENAKLKEIELEKOEILQAEI-ONGVKQIAPFS 82

Db 498 IKDVSLLIKRK-----RQROLVREOEKKQESSLSKQOVQSSASQTGIKQLPSAS 550

QY 83 -GTPPLHNSMVSENVIOSTAVTTVSS 107

Db 551 TGIP-----TASTTSAS 562

RESULT 15

US-09-854-856-28

; Sequence 28, Application US/09854856

; Patent No. 6541252

; GENERAL INFORMATION:

; APPLICANT: Walke, D. Wade

; APPLICANT: Hilbun, Erin

; APPLICANT: Donoho, Gregory

; TITLE OF INVENTION: Turner, C. Alexander Jr.
TITLE OF INVENTION: Encoding the Same

FILE REFERENCE: LEX-0178-USA

CURRENT APPLICATION NUMBER: US/09/854,856

CURRENT FILING DATE: 2001-05-14

PRIOR APPLICATION NUMBER: US 60/206,015

PRIOR FILING DATE: 2000-05-19

NUMBER OF SEQ ID NOS: 64

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 28

LENGTH: 801

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: VARIANT

LOCATION: (1)...(801)

OTHER INFORMATION: Xaa = Any Amino Acid

US-09-854-856-28

Query Match

Best Local Similarity 16.3%; Score 84; DB 4; Length 801;

Matches 30; Conservative 13; Mismatches 19; Indels 24; Gaps 5;

QY 25 LKQOVSLLKEKAILQATLREKKL-RVENAKLKEIELEKOEILQAEI-ONGVKQIAPFS 82

Db 558 IKDVSLLIKRK-----RQROLVREOEKKQESSLSKQOVQSSASQTGIKQLPSAS 610

QY 83 -GTPPLHNSMVSENVIOSTAVTTVSS 107

Db 611 TGIP-----TASTTSAS 622

Search completed: January 2, 2004, 18:23:41

Job time : 13.2824 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:22:28 ; Search time 24.1412 Seconds
(without alignments)
897.605 Million cell updates/sec

Title: US-09-930-169-2

Perfect score: 516
Sequence: 1 MANDAVLKRLEQKGAADQ.....NSMSENVIQSTAVTVSSG 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 733937

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PC1_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10C_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	516	100.0	108	US-09-930-169-2	Sequence 2, Appli
2	516	100.0	147	US-09-930-169-1	Sequence 1, Appli
3	511	99.0	312	US-09-851-026-36	Sequence 36, Appli
4	395	76.6	310	US-09-851-026-35	Sequence 35, Appli
5	93	18.0	1130	US-10-369-493-6751	Sequence 6751, Ap
6	92	17.8	1169	US-10-369-493-1095	Sequence 1095, Ap
7	90	17.4	956	US-10-080-608A-17	Sequence 17, Appli
8	90	17.4	956	US-10-370-685-106	Sequence 106, Appli
9	90	17.4	957	US-10-080-608A-16	Sequence 16, Appli
10	90	17.4	957	US-10-370-685-105	Sequence 105, Appli
11	89.5	17.3	815	US-10-080-608A-18	Sequence 18, Appli
12	89.5	17.3	815	US-10-370-685-107	Sequence 107, Appli
13	89	17.2	856	US-09-815-242-11310	Sequence 11310, A
14	89	17.2	1177	US-10-369-493-21558	Sequence 21558, A
15	87.5	17.0	1079	US-10-369-493-2116	Sequence 2116, Ap

16	85	16.5	223	US-09-925-300-1616	Sequence 1616, Ap
17	85	16.5	373	US-10-104-047-1995	Sequence 1995, Ap
18	84.5	16.4	857	US-10-369-493-9374	Sequence 9374, Ap
19	84.5	16.4	861	US-10-310-154-485	Sequence 485, App
20	84.5	16.4	861	US-10-369-493-17514	Sequence 17514, A
21	84.5	16.4	1526	US-10-369-493-22704	Sequence 22704, A
22	84.5	16.4	1965	US-10-369-493-3279	Sequence 3279, Ap
23	84.5	16.4	2288	US-10-369-493-6774	Sequence 6774, Ap
24	84	16.3	336	US-09-976-782-33	Sequence 33, Appli
25	84	16.3	2135	US-10-288-798-9	Sequence 9, Appli
26	84	16.3	2382	US-10-196-935A-2	Sequence 2, Appli
27	83.5	16.2	1884	US-09-785-770A-17	Sequence 17, Appli
28	83.5	16.2	1907	US-09-785-770A-16	Sequence 16, Appli
29	83	16.1	676	US-10-256-250-16	Sequence 16, Appli
30	83	16.1	724	US-10-023-634-80	Sequence 80, Appli
31	83	16.1	725	US-09-978-009A-47	Sequence 47, Appli
32	83	16.1	725	US-10-256-250-14	Sequence 14, Appli
33	83	16.1	725	US-10-023-634-78	Sequence 78, Appli
34	83	16.1	725	US-10-023-634-79	Sequence 79, Appli
35	83	16.1	1069	US-10-146-473-77	Sequence 77, Appli
36	83	16.1	1827	US-10-369-493-5368	Sequence 5368, Ap
37	82.5	16.0	210	US-10-154-251-67	Sequence 67, Appli
38	82.5	16.0	975	US-10-094-749-2035	Sequence 2035, Ap
39	82.5	16.0	1164	US-10-369-493-9770	Sequence 9770, Ap
40	82.5	16.0	1957	US-10-369-493-2070	Sequence 2070, Ap
41	82.5	16.0	2099	US-10-128-714-3290	Sequence 3290, Ap
42	82.5	16.0	2405	US-10-128-714-8290	Sequence 8290, Ap
43	82	15.9	166	US-09-930-169-3	Sequence 3, Appli
44	82	15.9	1188	US-10-369-493-17323	Sequence 17323, A
45	81.5	15.8	758	US-09-925-299-859	Sequence 859, App

ALIGNMENTS

RESULT 1
US-09-930-169-2
; Sequence 2, Application US/09930169
; Publication No. US20030004309A1
; GENERAL INFORMATION:
; APPLICANT: KIM, SUNGHON
; TITLE OF INVENTION: IMMUNOLOGICAL ENHANCEMENT AGENT COMPRISING N-TERMINAL PEPTIDE OF
; FILE REFERENCE: 058333/0106
; CURRENT FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: KR 2001-11310
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Mammalian
; OTHER INFORMATION: protein sequence
US-09-930-169-2

Query Match 100.0%; Score 516; DB 11; Length 108;
Best Local Similarity 100.0%; Pred. No. 2e-43;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MANDAVLKRLEQKGAADQIIIEYLKQVSLKKEKAILQATLREKKLRVENAKLKEIE 60
DB 1 MANDAVLKRLEQKGAADQIIIEYLKQVSLKKEKAILQATLREKKLRVENAKLKEIE 60
QY 61 ELKELIQAETIONGVKQIAFPSCGTPPLHANSWSENVIQSTAVTVSSG 108
DB 61 ELKELIQAETIONGVKQIAFPSCGTPPLHANSWSENVIQSTAVTVSSG 108

RESULT 2
US-09-930-169-1
Sequence 1, Application US/09930169
Publication No. US20030004309A1
GENERAL INFORMATION:
APPLICANT: KIM, SUNGHOON
APPLICANT: KO, YOUNG-GYU
TITLE OF INVENTION: IMMUNOLOGICAL ENHANCEMENT AGENT COMPRISING N-TERMINAL PEPTIDE OF
FILE REFERENCE: 058333/0106
CURRENT APPLICATION NUMBER: US/09/930,169
CURRENT FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: KR 2001-31310
PRIOR FILING DATE: 2001-06-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 147
TYPE: PRT
ORGANISM: Unknown Organism
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: Mammalian
US-09-930-169-1
Query Match 100.0%; Score 516; DB 11; Length 147;
Best Local Similarity 100.0%; Pred. No. 2,9e+43;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKKAIIQATLREKKLRVENAKLKEIE 60
DB 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKKAIIQATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIQAEIQNGVKQIAPFSGTPLHANSWSENVIOSTAVTVSSG 108
DB 61 ELKQELIQAEIQNGVKQIAPFSGTPLHANSWSENVIOSTAVTVSSG 108
RESULT 3
US-09-851-026-36
Sequence 36, Application US/09851026
Patent No. US20020160957A1
GENERAL INFORMATION:
APPLICANT: Stern, David M.
Clauss, Matthias
Kao, Janet
Kayton, Mark
Libutti, Steven K
TITLE OF INVENTION: Endothelial Monocyte Activating
Polypeptide II: A Mediator Which Activates Host Response
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham, LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.30, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/851,026
FILING DATE: 07-May-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/360,821
FILING DATE: 08-OCT-96
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 41735-A-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 312 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-851-026-36
Query Match 99.0%; Score 511; DB 10; Length 312;
Best Local Similarity 99.1%; Pred. No. 2,4e+42;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKKAIIQATLREKKLRVENAKLKEIE 60
DB 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKKAIIQATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIQAEIQNGVKQIAPFSGTPLHANSWSENVIOSTAVTVSSG 108
DB 61 ELKQELIQAEIQNGVKQIAPFSGTPLHANSWSENVIOSTAVTVSSG 108
RESULT 4
US-09-851-026-35
Sequence 35, Application US/09851026
Patent No. US20020160957A1
GENERAL INFORMATION:
APPLICANT: Stern, David M.
Clauss, Matthias
Kao, Janet
Kayton, Mark
Libutti, Steven K
TITLE OF INVENTION: Endothelial Monocyte Activating
Polypeptide II: A Mediator Which Activates Host Response
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham, LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.30, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/851,026
FILING DATE: 07-May-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/360,821
FILING DATE: 08-OCT-96
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41735-A-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 35;
US-09-851-026-35

Query Match 76.6%; Score 395; DB 10; Length 310;
Best Local Similarity 77.6%; Pred. No. 6,1e-31;
Matches 83; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

Qy 1 MANDAVLKRLEQKGAADQIIEYLKQVSLKKEKALIQATLREKKLRVENAKLKEIE 60
Db 1 MATNDVAVLKRLEQKGAADQIIEYLKQVALLKKEKALIQATLREKKLRVENAKLKEIE 60
Qy 61 ELKQELIQAEIQNGVQKQIAFPGSTPLHANSVSENVIOSTAVTVSS 107
Db 61 ELKQELIQAEIQNGVQKQIAFPGSTPLQNTCTASBSVQSPSVATTAS 107

RESULT 5
US-10-369-493-6751

Sequence 6751, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:

APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 6751
LENGTH: 1130
TYPE: PRT
ORGANISM: Caenorhabditis elegans

US-10-369-493-6751

Query Match 18.0%; Score 93; DB 12; Length 1130;
Best Local Similarity 24.6%; Pred. No. 1.7;
Matches 29; Conservative 30; Mismatches 33; Indels 26; Gaps 4;

Qy 2 ANNDVAVLKRLEQKGAADQIIEYLKQVSLKKEKALIQATLREKK-----KLRENAKL 55
Db 691 SDKNLLEELSK---NKNIEHLKQELQALNEKISTETEKDSLEKTIQLEIDNSSK 746
Qy 56 KKEIEEL-----KQELIQAEIQNGVQKQIAFPGSTPLHANSVSENVIOS 99
Db 747 SDQLEKHLRVNDMLDQGTIKDELIVKNEE--IKTISAKTQALLESSTVESSTTKLAS 802

RESULT 6
US-10-369-493-1095

Sequence 1095, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 1095
LENGTH: 1169

TYPE: PRT
ORGANISM: Methanococcus jannaschii
US-10-369-493-1095

Query Match 17.8%; Score 92; DB 12; Length 1169;
Best Local Similarity 26.6%; Pred. No. 2.2;
Matches 25; Conservative 24; Mismatches 25; Indels 20; Gaps 4;

Qy 3 NNDVAVLKRLEQKGAADQIIEYLKQVSLKKEKALIQATLREKKLRVENAKLKEIE 58
Db 281 NN--IINELNEKNE--EVELHKSILEVEIENDKVDLSINELKQVEIENKKE 336
Qy 59 IEELKQELIQ-----AEIQNGVQKQIAF 80
Db 337 IKETQKTIENRDSIIEKQIKIEEKIKNLVY 370

RESULT 7
US-10-080-608A-17

Sequence 17, Application US/10080608A
Publication No. US20030198956A1
GENERAL INFORMATION:

APPLICANT: Makowski, Lee
APPLICANT: Hyman, Paul
APPLICANT: Williams, Mark
TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
FILE REFERENCE: 8471-010-999
CURRENT APPLICATION NUMBER: US/10/080,608A
CURRENT FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 180
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 956
TYPE: PRT
ORGANISM: Mus musculus

US-10-080-608A-17

Query Match 17.4%; Score 90; DB 12; Length 956;
Best Local Similarity 26.1%; Pred. No. 2.7;
Matches 29; Conservative 25; Mismatches 43; Indels 14; Gaps 4;

Qy 6 AVLKRLKQKGAAD---QIIEYLKQVSLKKEKALIQATLREKK-----LRENAKL 55
Db 422 SLVRQLDDKQDEINQSQQLAKKLQ--MLDQBELASTRDYKIQEBLRLQIENEA 479
Qy 56 KKEIEELKQELIQAEIQNGVQKQIAFPGSTPLHANSVSENVIOSTAVTVSS 106
Db 480 KDEYKEVQLALELAVNYDQKQSEVEDKT--RANEQLTDELAQKTTTLTTT 528

RESULT 8
US-10-370-685-106

Sequence 106, Application US/10370685
Publication No. US20030215903A1
GENERAL INFORMATION:

APPLICANT: Hyman, Paul
APPLICANT: Goldberg, Edward
TITLE OF INVENTION: Nanostructures Containing PNA Joining and Functional Elements
FILE REFERENCE: NANF-P-004
CURRENT APPLICATION NUMBER: US/10/370,685
CURRENT FILING DATE: 2003-02-21
PRIOR APPLICATION NUMBER: 10/080,608
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 159
SOFTWARE: PatentIn version 3.2
SEQ ID NO 106
LENGTH: 956
TYPE: PRT
ORGANISM: mouse

US-10-370-685-106

Query Match 17.4%; Score 90; DB 12; Length 956;
Best Local Similarity 26.1%; Pred. No. 2.7;

```
Matches 29; Conservative 25; Mismatches 43; Indels 14; Gaps 4;
QY 6 AVLKRLQKGAAD--QIIEYLKQGVSLKEKAILQATLREKK-----LRVENAKL 55
DB 422 SLVRQDDKDEIQKSOQLAEKLRQ--MDDQDELLASTRDYEKIQEELTRIQIENEAA 479
QY 56 KKEIEELKQELLQAEIQNGVKQIAFPSSGTPHANSWSENVIOSTAVTVTS 106
DB 480 KQEVKEVLQALBELAVNYDQKSOEVEDKT--RANEQUTDELAKTTLTTLTT 528

RESULT 9
US-10-080-608A-16
; Sequence 16, Application US/10080608A
; Publication No. US20030198956A1
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/10/080,608A
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-080-608A-16

Query Match 17.4%; Score 90; DB 12; Length 957;
Best Local Similarity 26.1%; Pred. No. 2.7;
Matches 29; Conservative 25; Mismatches 43; Indels 14; Gaps 4;
QY 6 AVLKRLQKGAAD--QIIEYLKQGVSLKEKAILQATLREKK-----LRVENAKL 55
DB 423 SLVRQDDKDEIQKSOQLAEKLRQ--MDDQDELLASTRDYEKIQEELTRIQIENEAA 480
QY 56 KKEIEELKQELLQAEIQNGVKQIAFPSSGTPHANSWSENVIOSTAVTVTS 106
DB 481 KQEVKEVLQALBELAVNYDQKSOEVEDKT--RANEQUTDELAKTTLTTLTT 529

RESULT 10
US-10-370-685-105
; Sequence 105, Application US/10370685
; Publication No. US20030215903A1
; GENERAL INFORMATION:
; APPLICANT: Hyman, Paul
; APPLICANT: Goldberg, Edward
; TITLE OF INVENTION: Nanostructures Containing PNA Joining and Functional Elements
; FILE REFERENCE: NANF-P-004
; CURRENT APPLICATION NUMBER: US/10/370,685
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: 10/080,608
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 105
; LENGTH: 957
; TYPE: PRT
; ORGANISM: human
US-10-370-685-105

Query Match 17.4%; Score 90; DB 12; Length 957;
Best Local Similarity 26.1%; Pred. No. 2.7;
Matches 29; Conservative 25; Mismatches 43; Indels 14; Gaps 4;
QY 6 AVLKRLQKGAAD--QIIEYLKQGVSLKEKAILQATLREKK-----LRVENAKL 55
DB 423 SLVRQDDKDEIQKSOQLAEKLRQ--MDDQDELLASTRDYEKIQEELTRIQIENEAA 480
```

```
QY 56 KKEIEELKQELLQAEIQNGVKQIAFPSSGTPHANSWSENVIOSTAVTVTS 106
DB 481 KQEVKEVLQALBELAVNYDQKSOEVEDKT--RANEQUTDELAKTTLTTLTT 529

RESULT 11
US-10-080-608A-18
; Sequence 18, Application US/10080608A
; Publication No. US20030198956A1
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/10/080,608A
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 815
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-080-608A-18

Query Match 17.3%; Score 89.5; DB 12; Length 815;
Best Local Similarity 36.0%; Pred. No. 2.4;
Matches 27; Conservative 22; Mismatches 17; Indels 9; Gaps 5;
QY 9 KRLQKGAADQI--IEYLKQGVSLKEKAILQATLRE--EKKLRVENAKLKEIEELKQ 64
DB 440 QQLDEKDEIQKSOQLAEKLRQOV--LLOEAL--GTMRENEELIRENNRFOKEADKQ 496
QY 65 ELIQAEIQNGVKQIA 79
DB 497 E--GKEMMTALEIA 509

RESULT 12
US-10-370-685-107
; Sequence 107, Application US/10370685
; Publication No. US20030215903A1
; GENERAL INFORMATION:
; APPLICANT: Hyman, Paul
; APPLICANT: Goldberg, Edward
; TITLE OF INVENTION: Nanostructures Containing PNA Joining and Functional Elements
; FILE REFERENCE: NANF-P-004
; CURRENT APPLICATION NUMBER: US/10/370,685
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: 10/080,608
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 107
; LENGTH: 815
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-370-685-107

Query Match 17.3%; Score 89.5; DB 12; Length 815;
Best Local Similarity 36.0%; Pred. No. 2.4;
Matches 27; Conservative 22; Mismatches 17; Indels 9; Gaps 5;
QY 9 KRLQKGAADQI--IEYLKQGVSLKEKAILQATLRE--EKKLRVENAKLKEIEELKQ 64
DB 440 QQLDEKDEIQKSOQLAEKLRQOV--LLOEAL--GTMRENEELIRENNRFOKEADKQ 496
QY 65 ELIQAEIQNGVKQIA 79
DB 497 E--GKEMMTALEIA 509

RESULT 13
```

```
US-09-815-242-11310
; Sequence 11310, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyckind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11310
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11310

Query Match          17.2%; Score 89; DB 9; Length 856;
Best Local Similarity 38.2%; Pred. No. 2.9;
Matches 21; Conservative 14; Mismatches 20; Indels 0; Gaps 0;

Cy      11 LEQGAEDQIIEYLKQVSLKKEKALIQATLREKKLRVENAKLKKKEIEELKOE 65
Db      428 MEKESNAKQOEILKEISLKEKIQLEAFENEKEVFKETISRLKQWESLKEK 482

RESULT 14
US-10-369-493-21558
; Sequence 21558, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21558
; LENGTH: 1177
; TYPE: PRT
; ORGANISM: Pyrococcus abyssi
US-10-369-493-21558

Query Match          17.2%; Score 89; DB 12; Length 1177;
```

```
Best Local Similarity 29.3%; Pred. No. 4.4;
Matches 29; Conservative 20; Mismatches 24; Indels 26; Gaps 3;

Cy      4 NDAVLKRLKQGAADQIIEYLK-----QVSLKKE-----ALIQ 40
Db      408 NEADIKRLKLEKKEKRLSSRTITLKNKLPQIREVEKLEKKEKKAELSNVENKISSISQR 467

Cy      41 TLREKKLRVENAKLKK---EIEELKQELIQAEIQNGVK 76
Db      468 RRVKEELKKTSELQVSSLSLESLERELIKAEQSEVR 506

RESULT 15
US-10-369-493-2116
; Sequence 2116, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 2116
; LENGTH: 1079
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-2116

Query Match          17.0%; Score 87.5; DB 12; Length 1079;
Best Local Similarity 21.2%; Pred. No. 5.5;
Matches 29; Conservative 34; Mismatches 41; Indels 33; Gaps 4;

Cy      3 NNDVLRKRLKQGAADQIIEYLKQVSLKKEKALIQATLREKKLRVENA 53
Db      246 NVTALQWLEBKRRAREEQRIRREEARIAEBEKRLABVEARKERARLKKEKERKKE 305

Cy      54 KLR-----KEIEELKQELIQAEIQNGVKQIAFPSC---TPIHAN----- 89
Db      306 EMKQGGKRLSKKQEQALQRRLOQWLESGVRVAGLSNGKKQKPYTNKKKSNRSQTS 365

Cy      90 SMVESENVIOSTAVTVTS 106
Db      366 SISSGILBSSPATIS 382

Search completed: January 2, 2004, 18:31:22
Job time : 25.1412 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:20:38 ; Search time 141.882 Seconds
(without alignments)
692.625 Million cell updates/sec

Title: US-09-930-169-2
Perfect score: 516
Sequence: 1 MANDAVLKRLKRGKGAEDQ.....NSMVSENVIGSTAVTVSSG 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 5728757 seqs, 909918778 residues
Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_Main:*

1:	/cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
2:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*
3:	/cgn2_6/ptodata/1/paa/US07_COMB.pep.*
4:	/cgn2_6/ptodata/1/paa/US081_COMB.pep.*
5:	/cgn2_6/ptodata/1/paa/US082_COMB.pep.*
6:	/cgn2_6/ptodata/1/paa/US083_COMB.pep.*
7:	/cgn2_6/ptodata/1/paa/US084_COMB.pep.*
8:	/cgn2_6/ptodata/1/paa/US085_COMB.pep.*
9:	/cgn2_6/ptodata/1/paa/US086_COMB.pep.*
10:	/cgn2_6/ptodata/1/paa/US087_COMB.pep.*
11:	/cgn2_6/ptodata/1/paa/US088_COMB.pep.*
12:	/cgn2_6/ptodata/1/paa/US089_COMB.pep.*
13:	/cgn2_6/ptodata/1/paa/US090_COMB.pep.*
14:	/cgn2_6/ptodata/1/paa/US091_COMB.pep.*
15:	/cgn2_6/ptodata/1/paa/US092_COMB.pep.*
16:	/cgn2_6/ptodata/1/paa/US093_COMB.pep.*
17:	/cgn2_6/ptodata/1/paa/US094_COMB.pep.*
18:	/cgn2_6/ptodata/1/paa/US095_COMB.pep.*
19:	/cgn2_6/ptodata/1/paa/US096_COMB.pep.*
20:	/cgn2_6/ptodata/1/paa/US097_COMB.pep.*
21:	/cgn2_6/ptodata/1/paa/US098_COMB.pep.*
22:	/cgn2_6/ptodata/1/paa/US099_COMB.pep.*
23:	/cgn2_6/ptodata/1/paa/US100_COMB.pep.*
24:	/cgn2_6/ptodata/1/paa/US101_COMB.pep.*
25:	/cgn2_6/ptodata/1/paa/US102_COMB.pep.*
26:	/cgn2_6/ptodata/1/paa/US103_COMB.pep.*
27:	/cgn2_6/ptodata/1/paa/US104_COMB.pep.*
28:	/cgn2_6/ptodata/1/paa/US105_COMB.pep.*
29:	/cgn2_6/ptodata/1/paa/US106_COMB.pep.*
30:	/cgn2_6/ptodata/1/paa/US107_COMB.pep.*
31:	/cgn2_6/ptodata/1/paa/US108_COMB.pep.*
32:	/cgn2_6/ptodata/1/paa/US109_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	516	100.0	108	24	US-09-930-169-2
					Sequence 2, Appli

2	516	100.0	147	24	US-09-930-169-1	Sequence 1, Appli
3	516	100.0	312	1	PCT-US02-35563-47	Sequence 47, Appli
4	516	100.0	312	26	US-10-049-065-2	Sequence 2, Appli
5	511	99.0	151	23	US-09-834-366-16965	Sequence 16965, A
6	511	99.0	151	32	US-60-197-873-16965	Sequence 16965, A
7	511	99.0	203	32	US-60-453-050-11272	Sequence 11272, A
8	511	99.0	203	32	US-60-453-135-11272	Sequence 11272, A
9	511	99.0	203	32	US-60-453-444-5918	Sequence 5918, A
10	511	99.0	203	32	US-60-465-241-5918	Sequence 5919, Ap
11	511	99.0	203	32	US-60-466-412-11272	Sequence 11272, A
12	511	99.0	310	22	US-09-791-537-80174	Sequence 80174, A
13	511	99.0	312	23	US-09-851-026-36	Sequence 36, Appli
14	511	99.0	312	32	US-60-453-050-11270	Sequence 11270, A
15	511	99.0	312	32	US-60-453-135-11271	Sequence 11270, A
16	511	99.0	312	32	US-60-453-135-11271	Sequence 11271, A
17	511	99.0	312	32	US-60-455-444-5917	Sequence 5917, Ap
18	511	99.0	312	32	US-60-455-444-5918	Sequence 5918, Ap
19	511	99.0	312	32	US-60-465-241-5917	Sequence 5917, Ap
20	511	99.0	312	32	US-60-465-241-5918	Sequence 5918, Ap
21	511	99.0	312	32	US-60-466-412-11270	Sequence 11270, A
22	511	99.0	312	32	US-60-466-412-11271	Sequence 11271, A
23	511	99.0	312	32	US-09-949-016-11271	Sequence 11271, A
24	511	99.0	320	24	US-09-949-016-11271	Sequence 11271, A
25	511	99.0	341	22	US-09-760-475-2636	Sequence 2636, Ap
26	511	99.0	341	28	US-10-227-425-2636	Sequence 2636, Ap
27	425	82.4	359	22	US-09-791-537-12674	Sequence 12674, A
28	395	76.6	310	17	US-09-304-448-3	Sequence 3, Appli
29	395	76.6	310	22	US-09-791-537-30491	Sequence 30491, A
30	395	76.6	310	23	US-09-851-026-35	Sequence 35, Appli
31	106	20.5	294	20	US-09-614-150-4422	Sequence 4422, Ap
32	106	20.5	294	22	US-09-791-537-53480	Sequence 53480, A
33	106	20.5	294	32	US-60-167-217-4508	Sequence 4508, Ap
34	106	20.5	294	32	US-60-191-637-4437	Sequence 4437, Ap
35	100.5	19.5	672	19	US-09-537-043-22	Sequence 22, Appli
36	100.5	19.5	672	24	US-09-913-655-22	Sequence 22, Appli
37	96	18.6	163	22	US-09-791-537-90144	Sequence 90144, A
38	96	18.6	789	18	US-09-489-0394-9210	Sequence 9210, Ap
39	96	18.6	789	30	US-10-446-203-9210	Sequence 9210, Ap
40	96	18.6	1975	20	US-09-614-150-13074	Sequence 13074, A
41	96	18.6	1975	32	US-60-167-217-13120	Sequence 13120, A
42	96	18.6	1975	32	US-60-173-464-10670	Sequence 10670, A
43	96	18.6	1975	32	US-60-191-637-13115	Sequence 13115, A
44	96	18.6	1975	32	US-60-191-681-10316	Sequence 10316, A
45	96	18.6	2080	32	US-60-161-932-2022	Sequence 2022, Ap

ALIGNMENTS

RESULT 1
US-09-930-169-2
; Sequence 2, Application US/09930169
; GENERAL INFORMATION:
; APPLICANT: KIM, SUNGHOO
; APPLICANT: KO, YOUNG-GYU
; TITLE OF INVENTION: IMMUNOLOGICAL ENHANCEMENT AGENT COMPRISING N-TERMINAL PEPTIDE OF
; FILE REFERENCE: 058333/0106
; CURRENT FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: KR 2001-31310
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE: Description of Unknown Organism: Mammalian
; OTHER INFORMATION: Protein sequence
US-09-930-169-2

Query Match 100.0%; Score 516; DB 24; Length 108;
Best Local Similarity 100.0%; Pred. No. 2.2e-39;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANDAVLRLKLEQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIE 60
DB 1 MANDAVLRLKLEQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIOAEIONGVQKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108
DB 61 ELKQELIOAEIONGVQKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108

RESULT 2

US-09-930-169-1
Sequence 1, Application US/09930169

GENERAL INFORMATION:
APPLICANT: KIM, SUNGHON
TITLE OF INVENTION: IMMUNOLOGICAL ENHANCEMENT AGENT COMPRISING N-TERMINAL PEPTIDE OF
FILE REFERENCE: 058333/0106
CURRENT APPLICATION NUMBER: US/09/930,169
PRIOR FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: KR 2001-31310
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 1
LENGTH: 147
TYPE: PRT
ORGANISM: Unknown Organism
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: Mammalian
OTHER INFORMATION: protein sequence
US-09-930-169-1

Query Match 100.0%; Score 516; DB 24; Length 147;
Best Local Similarity 100.0%; Pred. No. 3.4e-39;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANDAVLRLKLEQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIE 60
DB 1 MANDAVLRLKLEQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIOAEIONGVQKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108
DB 61 ELKQELIOAEIONGVQKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108

RESULT 3

PCT-US02-35563-47
Sequence 47, Application PC/TUS0235563

GENERAL INFORMATION:
APPLICANT: Origene Technologies, Inc
TITLE OF INVENTION: REGULATED PROSTATE CANCER GENES
FILE REFERENCE: OGT 9U 803 PCT
CURRENT APPLICATION NUMBER: PCT/US02/35563
PRIOR FILING DATE: 2002-11-07
PRIOR APPLICATION NUMBER: 60/331,042
PRIOR FILING DATE: 2001-11-07
PRIOR APPLICATION NUMBER: 60/331,041
PRIOR FILING DATE: 2001-11-07
PRIOR APPLICATION NUMBER: 60/340,251
PRIOR FILING DATE: 2001-12-18
PRIOR APPLICATION NUMBER: 60/344,791
PRIOR FILING DATE: 2002-01-07
NUMBER OF SEQ ID NOS: 52
SOFTWARE: Patentin version 3.1
SEQ ID NO 47
LENGTH: 312
TYPE: PRT
ORGANISM: Homo sapiens

PCT-US02-35563-47

Query Match 100.0%; Score 516; DB 1; Length 312;
Best Local Similarity 100.0%; Pred. No. 9.7e-39;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANDAVLRLKLEQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIE 60
DB 1 MANDAVLRLKLEQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIOAEIONGVQKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108
DB 61 ELKQELIOAEIONGVQKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108

RESULT 4

US-10-049-065-2
Sequence 2, Application US/10049065

GENERAL INFORMATION:
APPLICANT: KIM, SUNGHON
APPLICANT: KO, YOUNG-GYU
APPLICANT: KIM, YOUNG SOO
TITLE OF INVENTION: P43 ANTI-TUMOR THERAPEUTIC AGENT AND THREE DIMENSIONAL
FILE REFERENCE: 058333/0113
CURRENT APPLICATION NUMBER: US/10/049,065
PRIOR FILING DATE: 2002-07-08
PRIOR APPLICATION NUMBER: PCT/KR00/00630
PRIOR FILING DATE: 2000-06-14
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 2
LENGTH: 312
TYPE: PRT
ORGANISM: Homo sapiens
US-10-049-065-2

Query Match 100.0%; Score 516; DB 26; Length 312;
Best Local Similarity 100.0%; Pred. No. 9.7e-39;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANDAVLRLKLEQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIE 60
DB 1 MANDAVLRLKLEQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIOAEIONGVQKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108
DB 61 ELKQELIOAEIONGVQKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108

RESULT 5

US-09-834-366-16965
Sequence 16965, Application US/09834366

GENERAL INFORMATION:
APPLICANT: Benjamin, Stephanie
APPLICANT: Tanaka, Hiroaki
APPLICANT: Dumas Malne Edwards, Jean Baptiste
APPLICANT: Jobert, Severin
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: 81 US2, REG
CURRENT APPLICATION NUMBER: US/09/834,366
PRIOR FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: US 60/197,873
PRIOR FILING DATE: 2000-04-18
NUMBER OF SEQ ID NOS: 52153
SOFTWARE: Patent .pm
SEQ ID NO 16965
LENGTH: 151
TYPE: PRT
ORGANISM: Homo sapiens
US-09-834-366-16965

Query Match 99.0%; Score 511; DB 23; Length 151;
Best Local Similarity 99.1%; Pred. No. 1e-38;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MANDDAVLRLEQKGAEDQIIEYLKQOVSLLKEKAILQATLRREKKLRVENAKLKEIE 60
Db 25 MANDDAVLRLEQKGAEDQIIEYLKQOVSLLKEKAILQATLRREKKLRVENAKLKEIE 84

Qy 61 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 108
Db 85 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 132

RESULT 6
US-60-197-873-16965
Sequence 16965, Application US/60197873

GENERAL INFORMATION:
APPLICANT: Bejani, Stephane
APPLICANT: Tanaka, Hiroaki
APPLICANT: Dumas Milne Edwards, Jean Baptiste
APPLICANT: Jobert, Severin
APPLICANT: Giordano, Jean-Yves
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: 81.US1.PRO
CURRENT APPLICATION NUMBER: US/60/197,873
CURRENT FILING DATE: 2000-04-18
NUMBER OF SEQ ID NOS: 52153
SOFTWARE: Patent.pm
SEQ ID NO 16965
LENGTH: 151
TYPE: PRT
ORGANISM: Homo sapiens
US-60-197-873-16965

Query Match 99.0%; Score 511; DB 32; Length 151;
Best Local Similarity 99.1%; Pred. No. 1e-38;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MANDDAVLRLEQKGAEDQIIEYLKQOVSLLKEKAILQATLRREKKLRVENAKLKEIE 60
Db 25 MANDDAVLRLEQKGAEDQIIEYLKQOVSLLKEKAILQATLRREKKLRVENAKLKEIE 84

Qy 61 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 108
Db 85 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 132

RESULT 7
US-60-453-050-11272
Sequence 11272, Application US/60453050

GENERAL INFORMATION:
APPLICANT: CARGILL, Michele
APPLICANT: LUKE, May
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
FILE REFERENCE: CLO01457
CURRENT APPLICATION NUMBER: US/60/453,050
CURRENT FILING DATE: 2003-03-10
NUMBER OF SEQ ID NOS: 82762
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11272
LENGTH: 203
TYPE: PRT
ORGANISM: Homo sapiens
US-60-453-050-11272

Query Match 99.0%; Score 511; DB 32; Length 203;
Best Local Similarity 99.1%; Pred. No. 1.5e-38;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MANDDAVLRLEQKGAEDQIIEYLKQOVSLLKEKAILQATLRREKKLRVENAKLKEIE 60
Db 61 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 108

Db 1 MANDDAVLRLEQKGAEDQIIEYLKQOVSLLKEKAILQATLRREKKLRVENAKLKEIE 60
Qy 61 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 108
Db 61 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 108

RESULT 8
US-60-453-135-11272
Sequence 11272, Application US/60453135

GENERAL INFORMATION:
APPLICANT: CARGILL, Michele
APPLICANT: IAKUBOVA, Olga
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
FILE REFERENCE: CLO01456
CURRENT APPLICATION NUMBER: US/60/453,135
CURRENT FILING DATE: 2003-03-10
NUMBER OF SEQ ID NOS: 82762
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11272
LENGTH: 203
TYPE: PRT
ORGANISM: Homo sapiens
US-60-453-135-11272

Query Match 99.0%; Score 511; DB 32; Length 203;
Best Local Similarity 99.1%; Pred. No. 1.5e-38;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MANDDAVLRLEQKGAEDQIIEYLKQOVSLLKEKAILQATLRREKKLRVENAKLKEIE 60
Db 1 MANDDAVLRLEQKGAEDQIIEYLKQOVSLLKEKAILQATLRREKKLRVENAKLKEIE 60

Qy 61 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 108
Db 61 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 108

RESULT 9
US-60-455-444-5919
Sequence 5919, Application US/60455444

GENERAL INFORMATION:
APPLICANT: CARGILL, Michele
APPLICANT: BEGOVICH, Ann
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
FILE REFERENCE: CLO01455
CURRENT APPLICATION NUMBER: US/60/455,444
CURRENT FILING DATE: 2003-03-18
NUMBER OF SEQ ID NOS: 50986
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5919
LENGTH: 203
TYPE: PRT
ORGANISM: Homo sapiens
US-60-455-444-5919

Query Match 99.0%; Score 511; DB 32; Length 203;
Best Local Similarity 99.1%; Pred. No. 1.5e-38;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MANDDAVLRLEQKGAEDQIIEYLKQOVSLLKEKAILQATLRREKKLRVENAKLKEIE 60
Db 1 MANDDAVLRLEQKGAEDQIIEYLKQOVSLLKEKAILQATLRREKKLRVENAKLKEIE 60

Qy 61 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 108
Db 61 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 108

RESULT 10
US-60-465-241-5919

; Sequence 5919, Application US/60465241
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: BEGOVICH, Ann
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001468
; CURRENT APPLICATION NUMBER: US/60/465,241
; CURRENT FILING DATE: 2003-04-23
; NUMBER OF SEQ ID NOS: 258418
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5919
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-465-241-5919

Query Match 99.0%; Score 511; DB 32; Length 203;
Best Local Similarity 99.1%; Pred. No. 1.5e-38;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKREKALLOATLREKKLRVENAKLKEIE 60
DB 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKREKALLOATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIOAEIONGVKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108
DB 61 ELKQELIOAEIONGVKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108

RESULT 11
US-60-466-412-11272
; Sequence 11272, Application US/60466412
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: IAKOUBOVA, Olga
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001466
; CURRENT APPLICATION NUMBER: US/60/466,412
; CURRENT FILING DATE: 2003-04-30
; NUMBER OF SEQ ID NOS: 429241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11272
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-466-412-11272

Query Match 99.0%; Score 511; DB 32; Length 203;
Best Local Similarity 99.1%; Pred. No. 1.5e-38;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKREKALLOATLREKKLRVENAKLKEIE 60
DB 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKREKALLOATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIOAEIONGVKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108
DB 61 ELKQELIOAEIONGVKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108

RESULT 12
US-09-791-537-80174
; Sequence 80174, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Biomomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBE
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537

; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 80174
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-537-80174

Query Match 99.0%; Score 511; DB 22; Length 310;
Best Local Similarity 99.1%; Pred. No. 2.8e-38;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKREKALLOATLREKKLRVENAKLKEIE 60
DB 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKREKALLOATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIOAEIONGVKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108
DB 61 ELKQELIOAEIONGVKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108

RESULT 13
US-09-851-026-36
; Sequence 36, Application US/09851026
; GENERAL INFORMATION:
; APPLICANT: Stern, David M.
; APPLICANT: Clause, Matthias
; Kao, Janet
; Kayton, Mark
; Libutti, Steven K
; TITLE OF INVENTION: Endothelial Monocyte Activating
; Polypeptide II: A Mediator Which Activates Host Response

NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham, LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA

ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.30, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/851,026
FILING DATE: 07-May-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/360,821
FILING DATE: 08-OCT-96

ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41735-A-PCT-US

TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525

INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 312 amino acids
TYPE: amino acid
STRANDEDNESS: single

MOLECULE TYPE: peptide
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 36:

US-09-851-026-36
Query Match 99.0%; Score 511; DB 23; Length 312;
Best Local Similarity 99.1%; Pred. No. 2.8e-38;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Search completed: January 2, 2004, 18:29:30
Job time : 142.882 secs

QY 1 MANDAVLKRLEQKGAADQIIEYIKQOVSLIKEKAILQATLREKKLRVENAKLKEIE 60
Db 1 MANDAVLKRLEQKGAADQIIEYIKQOVSLIKEKAILQATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIOAEIONGVKQIAPPSGTPLHANSWSENVIOSTAVTTVSSG 108
Db 61 ELKQELIOAEIONGVKQIAPPSGTPLHANSWSENVIOSTAVTTVSSG 108

RESULT 14

US-60-453-050-11270
; Sequence 11270, Application US/60453050
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: LUKE, May
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001457
; CURRENT APPLICATION NUMBER: US/60/453,050
; CURRENT FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11270
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-453-050-11270

Query Match 99.0%; Score 511; DB 32; Length 312;
Best Local Similarity 99.1%; Pred. No. 2.8e-38;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MANDAVLKRLEQKGAADQIIEYIKQOVSLIKEKAILQATLREKKLRVENAKLKEIE 60
Db 1 MANDAVLKRLEQKGAADQIIEYIKQOVSLIKEKAILQATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIOAEIONGVKQIAPPSGTPLHANSWSENVIOSTAVTTVSSG 108
Db 61 ELKQELIOAEIONGVKQIAPPSGTPLHANSWSENVIOSTAVTTVSSG 108

RESULT 15

US-60-453-050-11271
; Sequence 11271, Application US/60453050
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: LUKE, May
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001457
; CURRENT APPLICATION NUMBER: US/60/453,050
; CURRENT FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11271
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-453-050-11271

Query Match 99.0%; Score 511; DB 32; Length 312;
Best Local Similarity 99.1%; Pred. No. 2.8e-38;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MANDAVLKRLEQKGAADQIIEYIKQOVSLIKEKAILQATLREKKLRVENAKLKEIE 60
Db 1 MANDAVLKRLEQKGAADQIIEYIKQOVSLIKEKAILQATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIOAEIONGVKQIAPPSGTPLHANSWSENVIOSTAVTTVSSG 108
Db 61 ELKQELIOAEIONGVKQIAPPSGTPLHANSWSENVIOSTAVTTVSSG 108

THIS PAGE BLANK (USPTO)

PRIOR APPLICATION NUMBER: US 60/421,614
PRIOR FILING DATE: 2002-10-25
PRIOR APPLICATION NUMBER: US 60/421,552
PRIOR FILING DATE: 2002-10-25
PRIOR APPLICATION NUMBER: US 60/422,178
PRIOR FILING DATE: 2002-10-30
PRIOR APPLICATION NUMBER: US 60/422,177
PRIOR FILING DATE: 2002-10-30
PRIOR APPLICATION NUMBER: US 60/426,384
PRIOR FILING DATE: 2002-11-15
PRIOR APPLICATION NUMBER: US 60/426,394
PRIOR FILING DATE: 2002-11-15
PRIOR APPLICATION NUMBER: US 60/426,355
PRIOR FILING DATE: 2002-11-15
PRIOR APPLICATION NUMBER: US 60/426,430
PRIOR FILING DATE: 2002-11-15
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 2352
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1087
LENGTH: 123
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US03-33947-1087

Query Match Best Local Similarity 56.1%; Score 289.5; DB 1; Length 123;
Matches 66; Conservative 5; Mismatches 16; Indels 7; Gaps 3;

Qy 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKREKAILQATL--RESEKLRVENAKLKKE 58
Db 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKREKAILQATL--RESEKLRVENAKLKKE 58
Qy 59 IEELKQELIQAETONGVKQIAFPSCGTPPLHANSVSE 94
Db 57 V-ELKQELIQAETONGVKQIAFPSCGTPPLHANSVSE 91

RESULT 3
US-09-614-150A-4422
Sequence 4422, Application US/09614150A
GENERAL INFORMATION:
APPLICANT: Venter, J. Craig
TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
TITLE OF INVENTION: DROSOPHILA GENES.
FILE REFERENCE: CL000728
CURRENT APPLICATION NUMBER: US/09/614,150A
CURRENT FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/157,832
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: 60/160,191
PRIOR FILING DATE: 1999-10-19
PRIOR APPLICATION NUMBER: 60/161,932
PRIOR FILING DATE: 1999-10-28
PRIOR APPLICATION NUMBER: 60/164,769
PRIOR FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: 60/173,383
PRIOR FILING DATE: 1999-12-28
PRIOR APPLICATION NUMBER: 60/175,693
PRIOR FILING DATE: 2000-01-12
PRIOR APPLICATION NUMBER: 60/184,831
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/191,637
PRIOR FILING DATE: 2000-03-23
NUMBER OF SEQ ID NOS: 43008
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4422
LENGTH: 294
TYPE: PRT
ORGANISM: DROSOPHILA
US-09-614-150A-4422

Query Match Best Local Similarity 20.5%; Score 106; DB 5; Length 294;
Matches 24; Conservative 21; Mismatches 27; Indels 4; Gaps 1;

Qy 6 AVLKRLKQKGAADQIIIEYLKQVSLKREKAILQATLRESEKLRVENAKLKKEIEELKOE 65
Db 2 ADLQIASNNRERNALEISIGIQ-----QLVERQKQELIKENALAKEVEALAAQ 57
Qy 66 LQAETONGVKQIAFP 81
Db 58 LVQLELRNGKQI PVP 73

RESULT 4
US-09-614-150A-13074
Sequence 13074, Application US/09614150A
GENERAL INFORMATION:
APPLICANT: Venter, J. Craig
TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
TITLE OF INVENTION: DROSOPHILA GENES.
FILE REFERENCE: CL000728
CURRENT APPLICATION NUMBER: US/09/614,150A
CURRENT FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/157,832
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: 60/160,191
PRIOR FILING DATE: 1999-10-19
PRIOR APPLICATION NUMBER: 60/161,932
PRIOR FILING DATE: 1999-10-28
PRIOR APPLICATION NUMBER: 60/164,769
PRIOR FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: 60/173,383
PRIOR FILING DATE: 1999-12-28
PRIOR APPLICATION NUMBER: 60/175,693
PRIOR FILING DATE: 2000-01-12
PRIOR APPLICATION NUMBER: 60/184,831
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/191,637
PRIOR FILING DATE: 2000-03-23
NUMBER OF SEQ ID NOS: 43008
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13074
LENGTH: 1975
TYPE: PRT
ORGANISM: DROSOPHILA
US-09-614-150A-13074

Query Match Best Local Similarity 18.6%; Score 96; DB 5; Length 1975;
Matches 36; Conservative 15; Mismatches 41; Indels 46; Gaps 4;

Qy 11 LKQKGAADQIIIEYLKQVSLKREKAILQATLRESEKLRVENAKLKKEIEELKOE 37
Db 673 LKQKGAADQIIIEYLKQVSLKREKAILQATLRESEKLRVENAKLKKEIEELKOE 373
Qy 38 LQATLRESEKLRVENAKLKKEIEELKOE-----LQAETONGVKQIAFPSCGTPPLHANS 90
Db 739 LQATLRESEKLRVENAKLKKEIEELKOE-----LQAETONGVKQIAFPSCGTPPLHANS 90
Qy 91 MVENVIQSTAVTVSSG 108
Db 793 LKQKGAADQIIIEYLKQVSLKREKAILQATLRESEKLRVENAKLKKEIEELKOE 810

RESULT 5
PCT-US03-38193-1546
Sequence 1546, Application PC/TUS0338193
GENERAL INFORMATION:
APPLICANT: Aziz, Natasha
APPLICANT: Gineburg, Wendy M.


```

; APPLICANT: Zlotnick, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.00PC00
; CURRENT APPLICATION NUMBER: PCT/US03/38193
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1546
; LENGTH: 1046
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US03-38193-1546

```

```

Query Match      17.9%; Score 92.5; DB 1; Length 1046;
Best Local Similarity 26.9%; Pred. No. 10;
Matches 25; Conservative 30; Mismatches 31; Indels 7; Gaps 2;

```

```

Qy      8 LKRLQKGAADQIIEYLKQOVSLKKEKALIQATLREKKLRVENAKLKEIEELKQELI 67
Db      735 LKRLDVEYRGQAQAIIEFLKQISLAEEKML-----DYERLQRAEAQKQGVESLRKRL 788

```

```

Qy      68 QAEIQ-NGVKQIAFPSTPLHANSVSENVIOS 99
Db      789 VAENRLQAVBALCSSQHTMHIESNDISEETIRT 821

```

```

RESULT 6
US-10-723-860-1546
; Sequence 1546, Application US/10723860
; GENERAL INFORMATION:
; APPLICANT: Azitz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1546
; LENGTH: 1046
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-1546

```

```

Query Match      17.9%; Score 92.5; DB 6; Length 1046;
Best Local Similarity 26.9%; Pred. No. 10;
Matches 25; Conservative 30; Mismatches 31; Indels 7; Gaps 2;

```

```

Qy      8 LKRLQKGAADQIIEYLKQOVSLKKEKALIQATLREKKLRVENAKLKEIEELKQELI 67
Db      735 LKRLDVEYRGQAQAIIEFLKQISLAEEKML-----DYERLQRAEAQKQGVESLRKRL 788

```

```

Qy      68 QAEIQ-NGVKQIAFPSTPLHANSVSENVIOS 99
Db      789 VAENRLQAVBALCSSQHTMHIESNDISEETIRT 821

```

```

RESULT 7
US-10-258-899A-1169
; Sequence 1169, Application US/10258899A
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Dmanac, Radoje T.
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhou, Ping

```

```

; APPLICANT: Xu, Chongjun
; APPLICANT: Cao, Yicheng
; APPLICANT: Ma, Yungting
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Zhi Wei
; APPLICANT: Xue, Aidong
; APPLICANT: Yang, Yonghong
; APPLICANT: Wehrman, Tom
; APPLICANT: Goodrich, RyLe
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 787CIP2-2G/US
; CURRENT APPLICATION NUMBER: US/10/258,899A
; CURRENT FILING DATE: 2003-11-06
; PRIOR APPLICATION NUMBER: PCT/US01/04098
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 09/774,434
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: CnucCom
; SEQ ID NO 1169
; LENGTH: 1047
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-258-899A-1169

```

```

Query Match      17.9%; Score 92.5; DB 6; Length 1047;
Best Local Similarity 26.9%; Pred. No. 10;
Matches 25; Conservative 30; Mismatches 31; Indels 7; Gaps 2;

```

```

Qy      8 LKRLQKGAADQIIEYLKQOVSLKKEKALIQATLREKKLRVENAKLKEIEELKQELI 67
Db      736 LKRLDVEYRGQAQAIIEFLKQISLAEEKML-----DYERLQRAEAQKQGVESLRKRL 789

```

```

Qy      68 QAEIQ-NGVKQIAFPSTPLHANSVSENVIOS 99
Db      790 VAENRLQAVBALCSSQHTMHIESNDISEETIRT 822

```

```

RESULT 8
US-10-258-899A-3137
; Sequence 3137, Application US/10258899A
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Dmanac, Radoje T.
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhou, Ping
; APPLICANT: Xu, Chongjun
; APPLICANT: Cao, Yicheng
; APPLICANT: Ma, Yungting
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Dunrui

```

```
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Wang, Zhi Wei
APPLICANT: Xue, Aidong
APPLICANT: Yang, Yonghong
APPLICANT: Wehrman, Tom
APPLICANT: Goodrich, Ryle
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
FILE REFERENCE: 787CIP-26/US
CURRENT APPLICATION NUMBER: US/10/258, 899A
PRIOR FILING DATE: 2003-11-06
PRIOR APPLICATION NUMBER: PCT/US01/04098
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 09/774,434
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 09/728,422
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: 09/693,325
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 09/663,561
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 09/654,936
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 09/620,325
PRIOR FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/598,075
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 09/496,914
PRIOR FILING DATE: 2000-02-03
NUMBER OF SEQ ID NOS: 3960
SOFTWARE: Custom
SEQ ID NO 3137
LENGTH: 1047
TYPE: PRT
ORGANISM: Homo sapiens
US-10-258-899A-3137
```

```
Query Match 17.9%; Score 92.5; DB 6; Length 1047;
Best Local Similarity 26.9%; Pred. No. 10;
Matches 25; Conservative 30; Mismatches 31; Indels 7; Gaps 2;

QY 8 LKLEQKGAADQIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIELEKQELI 67
DB 736 LKLEQKGAADQIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIELEKQELI 67
QY 68 QAEIQ-NGVQKIAFPSTGPLHANSMSVENVIOSTAVTTVS 99
DB 790 VAEIRLQAEVLAELAVNVDQKSGVEVDKT--RANQQLDELQKTTTLTTT 822
```

```
RESULT 9
US-10-330-773-712
Sequence 712, Application US/10330773
GENERAL INFORMATION:
APPLICANT: David W. Morris
APPLICANT: Marc Malandro
TITLE OF INVENTION: Novel Compositions and Methods in Cancer
FILE REFERENCE: 529452001300
CURRENT APPLICATION NUMBER: US/10/330,773
CURRENT FILING DATE: 2002-12-27
NUMBER OF SEQ ID NOS: 981
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 712
LENGTH: 911
TYPE: PRT
ORGANISM: Homo sapiens
US-10-330-773-712
Query Match 17.4%; Score 90; DB 6; Length 911;
```

```
Best Local Similarity 26.1%; Pred. No. 14;
Matches 29; Conservative 25; Mismatches 43; Indels 14; Gaps 4;

QY 6 AVLKRLKQKGAAD---QIIEYLKQVSLKEKAILQATLREKK-----LRVENAKL 55
DB 379 SLVRLDQKDEINQSQSLAEKTKQ--MLDDELASTRRDYEKIQEELRLQLENEA 436
QY 56 KKEIELEKQELIQAEIQNGVQKIAFPSTGPLHANSMSVENVIOSTAVTTVS 106
DB 437 KQVEKVEVLAELAEVLAELAVNVDQKSGVEVDKT--RANQQLDELQKTTTLTTT 485
```

```
RESULT 10
US-10-330-773-709
Sequence 709, Application US/10330773
GENERAL INFORMATION:
APPLICANT: David W. Morris
APPLICANT: Marc Malandro
TITLE OF INVENTION: Novel Compositions and Methods in Cancer
FILE REFERENCE: 529452001300
CURRENT APPLICATION NUMBER: US/10/330,773
CURRENT FILING DATE: 2002-12-27
NUMBER OF SEQ ID NOS: 981
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 709
LENGTH: 952
TYPE: PRT
ORGANISM: Mus musculus
US-10-330-773-709
```

```
Query Match 17.4%; Score 90; DB 6; Length 952;
Best Local Similarity 26.1%; Pred. No. 15;
Matches 29; Conservative 25; Mismatches 43; Indels 14; Gaps 4;

QY 6 AVLKRLKQKGAAD---QIIEYLKQVSLKEKAILQATLREKK-----LRVENAKL 55
DB 420 SLVRLDQKDEINQSQSLAEKTKQ--MLDDELASTRRDYEKIQEELRLQLENEA 477
QY 56 KKEIELEKQELIQAEIQNGVQKIAFPSTGPLHANSMSVENVIOSTAVTTVS 106
DB 478 KQVEKVEVLAELAEVLAELAVNVDQKSGVEVDKT--RANQQLDELQKTTTLTTT 526
```

```
RESULT 11
US-10-679-063-2630
Sequence 2630, Application US/10679063
GENERAL INFORMATION:
APPLICANT: Edgeton, Michael D
TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
FILE REFERENCE: 38-15(52054)B
CURRENT APPLICATION NUMBER: US/10/679,063
CURRENT FILING DATE: 2003-10-02
PRIOR APPLICATION NUMBER: 60/415,758
PRIOR FILING DATE: 2002-10-02
NUMBER OF SEQ ID NOS: 27373
SEQ ID NO 2630
LENGTH: 300
TYPE: PRT
ORGANISM: Glycine max
US-10-679-063-2630
```

```
Query Match 17.3%; Score 89.5; DB 6; Length 300;
Best Local Similarity 26.7%; Pred. No. 5; 7;
Matches 28; Conservative 25; Mismatches 43; Indels 9; Gaps 2;

QY 2 ANNDVTLKLEQKGAADQIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIELE 61
DB 123 STNPADMRLKRRKVSNRD-----SARRSAREQQLSELQVEKLEKVENATLVKQPTD 176
QY 62 LKQELIQAEIQNGVQKIAFPSTGPLHANSMSVENVIOSTAVTTVS 106
DB 177 ASQHFREADTNRRVLKSDVEA--LRKVKLAEDMVTSSFTTLN 218
```

```

RESULT 12
US-09-614-150A-31653
Sequence 31653, Application US/09614150A
GENERAL INFORMATION:
APPLICANT: Venter, J. Craig
TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
TITLE OF INVENTION: DROSOPHILA GENES.
FILE REFERENCE: CL000728
CURRENT APPLICATION NUMBER: US/09/614.150A
CURRENT FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/157,832
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: 60/160,191
PRIOR FILING DATE: 1999-10-19
PRIOR APPLICATION NUMBER: 60/161,932
PRIOR FILING DATE: 1999-10-28
PRIOR APPLICATION NUMBER: 60/164,769
PRIOR FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: 60/173,383
PRIOR FILING DATE: 1999-12-28
PRIOR APPLICATION NUMBER: 60/175,693
PRIOR FILING DATE: 2000-01-12
PRIOR APPLICATION NUMBER: 60/184,831
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/191,637
PRIOR FILING DATE: 2000-03-23
NUMBER OF SEQ ID NOS: 43008
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 31653
LENGTH: 1305
TYPE: PRT
ORGANISM: DROSOPHILA
US-09-614-150A-31653

Query Match      16.9% Score 87; DB 5; Length 1305;
Best Local Similarity 30.2%; Pred. No. 33;
Matches 32; Conservative 22; Mismatches 38; Indels 14; Gaps 4

CY      2 ANDDAVLK-RLEQKGADQIIEYLKOQVSLIKERAILQATIRE-----EKKLRVEN 52
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
DB      37 SNOVVYAKRSRLAEIRIAQNDELIRGVKQLAQNKQLADLQRTAGCVTDLYGKEKQRIEL 96
       ||::||::||::||::||::||::||::||::||::||::||::||::||:

CY      53 AKTKREIE----LKQEL-IQAELONGVKQLAFFPGTLPHANMWS 93
       ||::||::||::||::||::||::||::||::||::||::||::||::||:
DB      97 EKRTRQIGERCGLKELDVGNCENLQEQLOLVAGLVEADVIS 142

RESULT 13
US-10-258-899A-3150
Sequence 3150, Application US/10258899A
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Drmanac, Radoje T.
APPLICANT: Asundi, Vinod
APPLICANT: Zhou, Ping
APPLICANT: Xu, Chongjun
APPLICANT: Cao, Yicheng
APPLICANT: Ma, Yunqiang
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyun
APPLICANT: Chen, Rui-hong
APPLICANT: Wang, Zhi Wei
APPLICANT: Xue, Aidong
APPLICANT: Yang, Yonghong
APPLICANT: Wehtman, Tom
APPLICANT: Goodrich, Ryle

```

```

; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 78CIP2-2G/US
; CURRENT APPLICATION NUMBER: US/10/258,899A
; PRIOR FILING DATE: 2003-11-06
; PRIOR APPLICATION NUMBER: PCT/US01/04098
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 09/774,434
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: CUSTOM
; SEQ ID NO 3150
; LENGTH: 931
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-258-899A-3150

Query Match      16.8%; Score 86.5; DB 6; Length 931;
Best Local Similarity 28.6%; Pred. No. 26;
Matches 28; Conservative 19; Mismatches 28; Indels 23; Gaps 4

Oy      3  NNDADVLRLEKGAEMADQIIYFLKQ--QVS-----LLEKKAIIQATLRRE---KULR 49
Db      476  NSEKQIKLEGGENHLNQTVSSLKQSQISAEARVKQILEKKNKILHESIKETSSKLSTIE 535
          |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy      50  VENAKKEI-----BELKQELIQAELONGVKQ 77
Db      536  FEKQIKKELEHYKEKGERADELENELHHEKENELLQ 573
          |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 14
; Sequence 1182, Application US/10258899A
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Drmanac, Radcoje T.
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhou, Ping
; APPLICANT: Xu, Chongjun
; APPLICANT: Cao, Yicheng
; APPLICANT: Ma, Yunqiang
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Duntui
; APPLICANT: Wang, Jlan-Rui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Zhi Wei
; APPLICANT: Xue, Aidong
; APPLICANT: Yang, Yonghong
; APPLICANT: Wehrman, Tom
; APPLICANT: Goodrich, Ryle
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 78CIP2-2G/US
; CURRENT APPLICATION NUMBER: US/10/258,899A
; CURRENT FILING DATE: 2003-11-06
; PRIOR APPLICATION NUMBER: PCT/US01/04098

```

